

# Evaluation of the Performance Evaluation Reform Act: Interim Report

## **Authors:**

Anthony Milanowski  
Jenna Aurand Scott  
Jackson Miller  
Matthew Finster  
*Westat*

Mimi Doll  
Holly Lewandowski  
Denise Roseland  
*Candeo Consulting*

Bradford R. White  
*Illinois Education Research Council*

Contributions also made by:  
Rula Zaru  
Clarissa McKithen  
*Westat*



# Contents

<u>Chapter</u>	<u>Page</u>
Executive Summary .....	viii
I. Introduction .....	1
II. Design and Methods.....	5
III. Document Review.....	11
Teacher Evaluation Systems.....	12
Principal Evaluation Systems .....	23
Concluding Remarks.....	32
IV Survey Data.....	33
Teacher Perceptions of the Evaluation Process.....	36
Principal and Assistant Principal Perceptions and Understanding of the Teacher Evaluation Process .....	53
Principal and Assistant Principal Perceptions of Their Own Evaluation Process .....	62
Principal Evaluator Perceptions of the Principal Evaluation Process .....	72
Additional Findings .....	80
V Case Study .....	86
Opportunities to Leverage PERA-Compliant Evaluation for Teacher and Principal Development .....	87
Teacher Evaluation .....	89
Principal Evaluation .....	90
Perceptions of System Validity .....	90
Teacher Evaluation .....	90
Principal Evaluation .....	94
Ongoing Communication and Training .....	96
Teacher Evaluation .....	96

Principal Evaluation.....	96
Trust/Collaboration.....	97
Teacher Evaluation.....	98
Principal Evaluation.....	99
Time/Effort/Skill to Complete Evaluation Cycle.....	100
Teacher Evaluation.....	101
Principal Evaluation.....	102
Early Impacts.....	102
Enhancing Ongoing Implementation .....	103
VI Conclusions .....	107
Research Questions.....	107
Question 1: How did districts evaluate teacher and principal professional practice? .....	108
Question 2: What are the perceptions of teachers, principals, and the evaluators about the evaluation systems? .....	112
Additional Findings.....	115
VII Recommendations .....	121
References.....	R-1

## Appendixes

A Document Review Protocol.....	A-1
B Document Review Template .....	A-1
C Detailed Case Study Methods.....	A-1
D Monitoring Protocol.....	A-1

# Executive Summary

Illinois recognizes the importance of strengthening its educator workforce by using evidence from educator evaluation results and best pedagogical practices to provide support to education professionals. To ensure that the state effectively strengthens the educator workforce, Illinois passed the Performance Evaluation Reform Act<sup>1</sup> (PERA) in 2010 to reform educator evaluation systems. Through PERA, Illinois aims to develop a highly effective educator workforce that provides students with a top-quality education.

The law requires that each Illinois district follow specified timelines to implement new teacher evaluation systems by the start of the 2016-17 school year (SY) and new principal evaluation systems by the beginning of SY 2012-13. The systems must:

1. Evaluate educators on a 4-point scale using multiple components (e.g., professional practice and student growth)
2. Deliver rigorous training to evaluators that requires obtaining certification
3. Develop professional development or remediation plans for ineffective educators

## Evaluation of the PERA-Compliant Educator Evaluation Systems

The Illinois State Board of Education (ISBE) contracted with Westat to conduct an evaluation of the PERA-compliant evaluation systems to assess: (1) the validity and reliability of the systems, (2) how the systems **inform staff development**, and (3) the systems' **relationship to student achievement growth**. The study began in January 2013 and will end in December 2015. Due to shifting implementation timelines, the interim report focuses on evaluating the PERA-compliant evaluation systems for the 2013-14 SY. The report primarily studies Race to the Top (RTT) districts and school improvement grant (SIG) recipients that have implemented or are in the process of implementing the evaluation systems. Since the PERA

---

<sup>1</sup> <http://www.ilga.gov/legislation/fulltext.asp?DocName=09600SB0315enr&GA=96&SessionId=76&DocTypeId=SB&LegID=&DocNum=0315&GAID=10&Session=>

evaluation systems rollout is still ongoing, the interim report addresses the following two overarching descriptive study questions that assess implementation progress:

1. What did districts do in SY 2013-14 to implement PERA-compliant teacher and principal performance evaluation systems?
2. What are the perceptions of teachers, principals, and the evaluators about the evaluation systems?

Using a mixed-methods research design that incorporates: (1) a document review, (2) multiple surveys, and (3) a case study, the report provides several important findings and recommendations.

### Summary Findings

The study findings show that most study districts are implementing PERA-compliant evaluation systems. Major findings by study question are listed below.

#### What did districts do in SY 2013-14 to implement PERA-compliant teacher and principal performance evaluation systems?

1. Teacher evaluation systems use a professional practice framework, including: (1) using rubrics with specified rating levels, (2) including the appropriate number of teacher observations, and (3) training evaluators. The majority of districts use the Danielson Framework for Teaching to evaluate teacher professional practice.
2. Most districts use a principal/assistant principal professional practice framework, including: (1) conducting two formal principal/assistant principal observations; (2) informing principals and assistant principals about evaluation requirements; (3) having principals, assistant principals, and their evaluators set principal and assistant principal performance goals; (4) requiring principals and assistant principals to conduct self-evaluations; (5) providing evaluation feedback to principals and assistant principals; and (6) offering mandatory evaluator training. The majority of districts use the Illinois Performance Standards for School Leaders rubric, although some used significantly different frameworks.
3. The majority of districts use or intend to use growth targets or goals to meet the student growth requirement for both teachers and principals/assistant principals. Most districts base the student growth rating on either the number of goals met or exceeded or an average of ratings made on each goal. While the student growth process is more developed in principal/assistant principal evaluation systems, fewer districts are as far along in defining the process for their teacher evaluation systems. In some instances, the selected growth measures appear difficult to use to make

comparisons across educators and schools (and certainly districts). Further, it appears that there are some data quality concerns in some growth measures due to less sophisticated data linkages.

4. Some districts are still piloting the student growth component of their teacher evaluation system and therefore have not yet determined how professional practice, growth, and other measures will combine for a summative evaluation rating. However, other districts have implemented student growth measures and are combining them with practice ratings.
5. Almost all districts have fully operational principal/assistant principal evaluation systems combining leadership practice and student growth measures.
6. Most districts have developed documentation and processes to communicate the new evaluation systems to educators. The processes seem more defined for teachers than principals/assistant principals. In addition, communication is weaker on documenting and describing student growth processes compared to professional practice.
7. The majority of districts have a process in place to train evaluators and have taken actions to promote inter-rater agreement. However, it was less common for districts to promote inter-rater agreement or consistency for principal evaluators.

#### What are the perceptions of teachers, principals, and the evaluators about the evaluation systems?

1. In most districts, the majority of educators and evaluators have a strong understanding of the professional practice rubrics, but not as strong an understanding of the student growth measures or how their districts combine the measures to determine a final evaluation rating.
2. In most districts, almost all educators have a more favorable perception of the professional practice component than the student growth component.
3. In most districts, the majority of educators believe that the evaluation results are useful for driving professional development and improving their practice, but some believe that districts have a difficult time truly linking the results to effective professional development.
4. In most districts, the majority of educators are optimistic that PERA-compliant evaluation systems will have positive impacts in the future.
5. In most districts, some educators are concerned that the evaluation systems may require more time and effort than they are worth.

## Summary Recommendations

Drawing on the study findings, the research team offers seven recommendations to further enhance the PERA evaluation systems' effectiveness.

- ▶ Disseminate more intensive technical assistance on developing student growth measures to districts. Despite existing guidance documents on growth measures, many districts need additional and/or more directive guidance.
- ▶ Specify more concrete rules on student growth and methods for addressing data and measurement quality to districts. In addition, ISBE may want to provide greater technical assistance on these issues by leveraging the support of regional universities or educational agencies.
- ▶ Develop more prescriptive guidance on measuring student growth using Types I, II, and III assessments to: (1) facilitate comparisons that are more valid across teachers and (2) maximize the utility of the three assessments. Suggested options include:
  - Require districts to use more standardized and uniform growth measurement models, such as value-added, student growth percentiles, or similar statistically motivated models for Types I/II assessments.
  - Develop a state-administered value-added or similar model that compares growth across the state. Districts could then be given the option of using results from this model in their teacher (and potentially also principal) evaluation systems instead of developing their own statistical model.
  - Develop a state-administered value-added or similar model for districts to use as a yardstick to assess the validity and rigor of local student growth measures.



- ▶ Revisit and clarify the distinction between growth measures based on Type I and Type II assessments. Despite definitions and examples in the PERA guidance, many districts are unclear about the differences between the assessments and their respective measures.
- ▶ Provide more guidance and technical assistance on achieving inter-rater agreement or consistency among teacher and principal/assistant principal raters. Suggestions include: (1) documenting and sharing district best practices, (2) describing and disseminating model rater calibration and re-certification programs,<sup>2</sup> and (3) encouraging districts to do co-observations and examine the rater agreement for a sample of those evaluated.
- ▶ Further develop the Illinois Performance Standards for School Leaders rubric and find ways to involve principals more closely in this process. In addition, the state rubrics may benefit from a streamlined approach that requires less work to apply and clarifies some of the standards.
- ▶ Ensure a stronger link between district evaluation results and professional development. Districts should systematically evaluate their professional development programs to ensure that opportunities directly related to the evaluation rubric performance dimensions are available and communicated to educators.

---

<sup>2</sup> For example, in order to help evaluators stay in calibration, some districts have evaluators meet periodically to view and rate video clips or discuss how they are interpreting the rubrics in the field. Some require evaluators to pass a recertification test after a substantial interval from the initial training, to ensure that evaluators have not drifted apart in how they are applying the rubrics.



# Introduction

*Recognizing the importance of strengthening the educator workforce to improve the education profession and provide students with high-quality education, Illinois passed the Performance Evaluation Reform Act<sup>3</sup> (PERA) in 2010. PERA requires that all Illinois school districts establish new educator evaluation systems. The systems must evaluate educators on a 4-point scale and include multiple components, including professional practice and student growth. In addition, districts must ensure that evaluators undergo rigorous training and obtain certification. As defined by PERA, districts must also develop professional development or remediation plans for ineffective teachers. PERA mandates that districts implement new principal evaluation systems by the beginning of the 2012-13 school year (SY). However, the law allows for staggered district implementation of new teacher evaluation systems. Notably, while all districts had to have incorporated certain new teacher evaluation components by the 2012-13 SY (such as a four-level rating system), all districts do not have to incorporate student growth until the start of the 2016-17 SY. PERA specifies that districts include student growth on a staggered schedule. Some districts, such as School Improvement Grant (SIG) recipients and certain Race to the Top (RTT) participants, have already implemented, piloted, or are starting to implement student growth as a component in their teacher evaluation systems.*

---

<sup>3</sup> <http://www.ilga.gov/legislation/fulltext.asp?DocName=09600SB0315enr&GA=96&SessionId=76&DocTypeId=SB&LegID=&DocNum=0315&GAID=10&Session=>

## **Independent Evaluation of the PERA-Compliant Educator Evaluation Systems**

In compliance with PERA, the Illinois State Board of Education (ISBE) contracted with Westat to conduct an evaluation of the PERA-compliant evaluation systems to assess: (1) the validity and reliability of the systems, (2) how the systems inform staff development, and (3) the systems relationship to student achievement growth. The study began in January 2013 and will end in December 2015. Due to shifting implementation timelines, the interim report focuses on evaluating the PERA-compliant evaluation systems for the 2013-14 SY. The report primarily studies RTT and SIG districts that have implemented or are in the process of implementing the evaluation systems. Since the PERA evaluation systems rollout is still ongoing, the interim report addresses the descriptive study questions, listed below, that assess implementation progress.

### **Study Questions**

1. What did districts do in SY 2013-14 to implement PERA-compliant teacher and principal performance evaluation systems?
  - How did districts evaluate teacher and principal professional practice?
  - How did districts measure student growth?
  - How did districts combine ratings of teacher and principal professional practice and student growth to determine an overall summative evaluation rating?
  - How did districts communicate about their evaluation systems to teachers and principals?
  - In what areas have districts made progress toward full implementation of PERA-compliant teacher and principal evaluation systems, and in what areas are there concerns or difficulties?

2. What are the perceptions of teachers, principals, and the evaluators about the evaluation systems?
  - Did teachers, principals, and evaluators understand the evaluation systems that districts piloted or used in the 2013-14 SY?
  - Did teachers and principals perceive these systems as fair?
  - Did teachers, principals, and evaluators perceive that the evaluations affected teacher and principal professional development?
  - Did teachers, principals, and evaluators perceive that the evaluations affected instruction or leadership practice?

To answer the study questions, the research team uses a mixed-methods research design that incorporates: (1) a document review, (2) multiple surveys, and (3) a case study. Individual chapters offer detailed descriptions of each method and its respective findings. The report is organized as follows:

- ▶ Chapter 1 – Introduction
- ▶ Chapter 2 – Design and Methods
- ▶ Chapter 3 – Document Review
- ▶ Chapter 4 – Surveys
- ▶ Chapter 5 – Case Study
- ▶ Chapter 6 –Conclusions
- ▶ Chapter 7 – Recommendations
- ▶ Appendixes



# Design and Methods

*To most effectively provide ISBE with a comprehensive report on the rollout of its statewide PERA-compliant evaluation systems, the study team designed a study to answer the following research questions:*

1. What did districts do in SY 2013-14 to implement PERA-compliant teacher and principal performance evaluation systems?
  - How did districts evaluate teacher and principal professional practice?
  - How did districts measure student growth?
  - How did districts combine ratings of teacher and principal professional practice and student growth to determine an overall summative evaluation rating?
  - How did districts communicate about their evaluation systems to teachers and principals?
  - In what areas have districts made progress toward full implementation of PERA-compliant teacher and principal evaluation systems, and in what areas are there concerns or difficulties?
2. What are the perceptions of teachers, principals, and the evaluators about the evaluation systems?
  - Did teachers, principals, and evaluators understand the evaluation systems that districts piloted or used in the 2013-14 SY?
  - Did teachers and principals perceive these systems as fair?
  - Did teachers, principals, and evaluators perceive that the evaluations affected teacher and principal professional development?
  - Did teachers, principals, and evaluators perceive that the evaluations affected instruction or leadership practice?

In addressing the research question, the study team used a mixed-method study design, drawing on both quantitative and qualitative data collected through (1) a document review; (2) teacher, principal, and evaluator surveys; and (3) a case study. A description of the study sample is below, followed by a comprehensive description of each method.

### **Sample of Districts Included in the Study**

In collaboration with ISBE, in October 2013 the research team contacted and requested the participation of the 35 RTT and SIG districts that implemented or piloted a PERA-compliant teacher evaluation system in the 2013-14 SY. All requested districts participated in the study, with the exception of two RTT districts that dropped out of the RTT program during the 2013-14 SY. ISBE also asked additional districts implementing PERA-compliant principal evaluation systems to volunteer to participate in the study through a message in the monthly **ISBE Superintendents' Bulletin**. **As a result of this request, three additional districts volunteered to participate.**

The number of districts participating in the study varied across the three data collection methods. Thirty-two districts submitted information for document review. Teachers from all 35 study districts (meaning RTT) participated in the teacher survey; principals and assistant principals from 23 districts participated in the principal survey; other evaluators of teachers (meaning those other than principals and assistant principals) from 10 districts participated in the teacher evaluator survey; and superintendents, assistant superintendents, and other evaluators of principals from 25 districts participated in the principal evaluator survey. For the case study, the study team strategically and purposefully limited the sample to six<sup>4</sup> districts.

---

<sup>4</sup> However, only five districts are represented in the interim report due to interview schedule conflicts. The study team will complete these interviews in August 2014, and the results will inform a supplement to this report.



## Data Collection Methodology

### Document Review

To comprehensively examine the study districts' implementation of PERA-compliant systems, the study team followed a systematic process for collecting documentation on each district's evaluation system. First, the study team reviewed ISBE- and performance evaluation advisory panel (PEAC)-developed PERA regulations and guidance documents. Based on the review, the team developed a protocol for collecting information on each required component of a PERA-compliant evaluation system. After developing the protocol, the team requested information from districts (e.g., guidebooks/handbooks or links to district websites) describing their educator evaluation and policies. Thirty-two districts submitted documentation, including 28 RTT districts and 4 additional districts, including the 3 volunteer districts and 1 former RTT district that submitted documentation.<sup>5</sup> The team followed up with districts by email or phone either to clarify information from the documentation or to request additional information. Based on the documentation, the team used a template to develop district-level summaries that focused on practice evaluation, student growth measures, and the strategies used to combine the two measures into a final evaluation rating. The team then reviewed each summary that described key features of each district's evaluation system and tabulated the important features of the systems across districts.

### Surveys

The study team collected survey data from teachers, principals, assistant principals, and evaluators to assess perceptions of the fairness, complexity, and implementation of the evaluation systems. The team developed five sets of survey items covering themes relevant to teachers, principals, teacher evaluators, and principal evaluators. To reduce burden on principals and assistant principals, who are both evaluators of teachers and recipients of evaluations, they randomly received one of two surveys; one contained items on the principals' perceptions of their role as an evaluator, and the other assessed their perceptions of the principal evaluation system. Some of the survey items drew from surveys of evaluation system implementation other organizations conducted, including the Consortium on Chicago School Research (CCSR). The team drafted the remaining items to reflect specific PERA requirements.

---

<sup>5</sup> Because of the varying PERA roll-out timeline that specifies a different timeline for various subsets of districts, the four volunteer districts were included only in the principal evaluation documentation review. The Document Review chapter discusses this in more detail.

To administer the surveys to principals and evaluators, the study team requested email addresses in study districts for principals, assistant principals, third-party evaluators, and principal evaluators such as superintendents and district-level staff. The team received email addresses from 22 RTT districts and the 3 volunteer districts. In May 2014, the team sent email invitations to participate in a web-based survey to all administrators and evaluators for which it had an email address and followed up with reminders each week until the survey closed in June 2014. Noted above, respondents from 10 districts participated in the teacher evaluator survey; respondents from 21 districts participated in the principal survey on evaluating teachers; respondents from 23 districts participated in the principal survey on their own evaluation system; and respondents from 25 districts participated in the principal evaluator survey. The response rate for each of the surveys was

- ▶ 69 percent for the survey of teacher evaluators;
- ▶ 76 percent for the principal survey on evaluating teachers;
- ▶ 77 percent for the principal survey on the principal evaluation system; and
- ▶ 73 percent for the principal evaluator survey.

After collecting and cleaning the data, the study team ran a series of district-level descriptive analyses summarizing the survey results. Our analysis included basic frequencies and summary statistics for each item in the surveys. Additional detail about the analysis is included at the beginning of the chapter on the administrator and evaluator surveys.

In addition to the administrator and evaluator surveys, the team worked with CCSR and the University of Chicago Impact to survey teachers across Illinois in March and April of 2014, as part of the 5Essentials Survey. This collaboration included a review of proposed items on CCSR's prior surveys **and a paring down of the number of items on Westat's proposed** teacher survey to minimize the burden placed on teachers and promote a high response rate. Impact administered the 5Essential Survey to all districts except Chicago Public Schools (CPS), which included 44 questions pertaining to teacher evaluation systems. CCSR administered its own version of the survey with the same 44 questions to teachers in CPS. The overall response rate was 81 percent for CPS and 58 percent for the rest of the state. All 35 study districts had survey respondents. Analysis of the teacher survey data included frequencies and summary statistics for each item. In addition, the team used *t*-tests to examine differences in how teachers in study and non-study districts responded to specific items.

### Case Study

To develop a more in-depth assessment of how districts of varying sizes and locations implemented PERA in SY 2013-14, the study team selected five districts that began implementing PERA-compliant teacher evaluation systems. The team chose the six districts from the RTT grantees and volunteer districts in order to represent a cross-section of locations, student demographics, and sizes. The team conducted 61 interviews with principals, teachers, district administrators, union representatives, teacher evaluators, and principal evaluators. In addition, the team analyzed documentation that described the evaluation design and implementation process prior to interviews to focus interview questions on information that was not included in existing documentation. The study team analyzed data from interviews across and within districts using NVivo qualitative research software. Appendix C describes the methods used in this part of the study.



# Document Review

In order to describe the evaluation processes districts implemented to comply with PERA, the research team first requested that districts share any documentation they had about their teacher and principal evaluation processes. The team asked them to provide documents that described how they evaluate teaching and leadership practice, how they measure student growth, and how they combine the two to determine a final summative evaluation rating. To summarize **each district's evaluation processes and to compare and contrast the processes** across districts, the study team developed a protocol (see Appendix A) and accompanying template (see Appendix B). The team used the district documents to fill in the template. After reviewing the documents, if there were outstanding questions about the teacher and/or principal evaluation processes, the team contacted the district to set up an interview with the study contact person to go over the documents and resolve these questions. If the team needed further clarification, members talked to additional district staff or followed up with email requests for additional information.

The study team received usable information from 32 districts, including 28 of the RTT districts, 1 former RTT district,<sup>6</sup> and 3 volunteer districts that were implementing PERA-compliant evaluation systems for principals and assistant principals. The amount of information obtained from each district varied. Team members were able to talk with staff knowledgeable about the evaluation system in 30 of the districts, but some of these districts **were unable to provide answers to all of the study's questions.**

This section of the report describes findings from the review, disaggregated by teacher and principal evaluation systems. Described earlier in the report, PERA requirements mandate that both evaluation systems include two components (or inputs), teaching or leadership professional practice (practice) and student growth (growth), in determining teacher and principal evaluation ratings. Therefore, the teacher and principal evaluation systems sections describe (in respective order): (1) the practice component of the evaluation, (2) the growth component of the evaluation, and (3) the aggregate (or summative) evaluation score. The teacher evaluation system findings report only information received from RTT districts, since

---

<sup>6</sup> The district submitted information to the study team during the process of dropping out of RTT.

the other districts were not required under PERA to have PERA-compliant teacher evaluation systems for the 2013-14 SY.<sup>7</sup>

## Teacher Evaluation Systems

### Evaluating Teaching Practice

Per PERA requirements, districts must evaluate teachers on their practice. To evaluate the quality of teacher practice, district evaluators use a rating scale or rubric to observe teachers in action (e.g., teaching a lesson) and review artifacts (e.g., lesson plans, student work).

### *Practice Rubrics, Observations, and Additional Evidence Districts Used*

Twenty-seven of the RTT districts reported using the Danielson Framework for Teaching<sup>8</sup> (the Framework) or a lightly modified version of the Framework as the basis for evaluating teaching practice. Most of the modifications entailed a reduction in the number of rubric components. All of these rubrics defined four performance levels, as PERA requires. The remaining district used a tool called the Teaching Essentials that the district modified to align with the Illinois teaching standards. This tool differs substantially from the Framework in that it does not appear to have rubrics that define the four performance levels. More than half of the RTT districts (16) developed modified or specialized versions of their rubrics for non-classroom certified staff (e.g., librarians/media specialists, counselors), and one district was in the process of developing versions for these staff.

Observations of teachers are a key source of evidence for rating teaching practice. Almost all of the districts reported that tenured teachers received or will receive one formal observation every other year. Most districts also did one informal observation of tenured teachers. One district planned to conduct an informal observation in the years between formal observations. Non-tenured teachers received two observations per year, as PERA requires, in all districts. In five districts, non-tenured teachers received three formal observations. The districts varied considerably in the frequency of informal observations, with the typical number being one to three, although a few districts reported no limit on the number (e.g., allowing principals to make the determination). Further, all but one district reported that the length of the formal observations was one class period, typically 40-45 minutes. One district allowed for several shorter periods of observation to be combined to equal one class period. Informal observations were typically 5-10 minutes.

---

<sup>7</sup> SIG districts were also required to implement in SIG schools. However, all of the SIG districts that provided information on their teacher evaluation systems were also RTT districts.

<sup>8</sup> <http://danielsongroup.org/framework/>

Districts varied widely in how they recorded information during observations. In some districts, evaluators scripted the lesson they observed and then made ratings later after reviewing the script. Others had the evaluators observe with the rubric form and take notes relevant to each standard or component on the form. Some districts left the method of recording up to the observer.

To ensure a more complete picture of teaching practice, all but three districts required the collection of artifacts such as lesson plans, student work, professional development logs, parent communications, or sample assessments to supplement the observations. The other three districts allowed but did not require the collection of artifacts. Several of the districts provided evaluators with a list of example artifacts that they could use as evidence in making an evaluation rating.

### *Evaluator Training*

Training of evaluators is critical to ensure quality teacher practice ratings. Trained raters are more likely to apply the practice rubrics accurately and consistently. All districts reported that their evaluators completed the state-required training provided by Growth Through Learning.<sup>9</sup> Eighteen of the districts provided additional training to evaluators. The length of additional training varied widely, and some districts were not able to report the number of training hours. Of those that provided training hours, the range was from 5 to 40 hours. Districts obtained additional training from a variety of sources, including Regional Offices of Education, the Danielson group, the American Institutes for Research, the Consortium for Educational Change, and district personnel.

### *Inter-Rater Agreement*

One desirable result of evaluator training is consistent application of the rubrics from evaluator to evaluator. Inter-rater agreement is one indicator of this consistency, measuring the degree to which different raters using the same rubric give the same rating in identical situations. Inter-rater agreement is important because teachers are likely to find evaluation results more credible if they are assured that their ratings would be similar if they were rated by a different evaluator. Thirteen districts reported efforts to assess the degree to which different raters were likely to agree on their ratings. Two of the districts calculated or planned to calculate some form of agreement statistic, but neither was able to report the results at the time of data collection. Eleven districts used a variety of informal methods to check on and promote rater agreement. The most common method, which five districts used, was to encourage or require evaluators to regularly discuss ratings and interpretations of the rubric. Three districts used co-observation. (In one of these districts, representatives of the local

---

<sup>9</sup> <http://www.growththroughlearningillinois.org/>

teachers' association observed alongside administrators.) Two districts mentioned annual recalibration training. In one district, evaluators rated and discussed videos of teaching at regular professional development sessions. However, 11 districts reported that they had not yet assessed inter-rater agreement. Two of these planned to train for agreement in the summer or fall of 2014. Two others (small districts) had only one evaluator. Five districts did not provide information on inter-rater agreement.

#### *Overall Rating of Teaching Practice*

To determine the aggregate teacher practice score, most districts (20) used a defined, mechanical process, after evaluators had used their judgment to rate each of the standards or components of practice. The two most common methods were: (1) to sum or average the one to four ratings on each component, then compare the result to a table that provided score ranges required for each on the four overall practice levels and (2) to require that a certain number of components be rated at each level (e.g., to be rated at the highest level, the teacher needs 13 components rated at the highest level and no more than 1 at the needs improvement or unsatisfactory level.) Seven districts, however, allowed the evaluator to use his/her own method to determine a final practice rating. Two districts were unable to provide enough information for the team to determine how the final practice rating was made.

#### *Peer Evaluation*

In addition to having administrators evaluate teacher practice, some districts used peer evaluators. Of the 28 districts providing information, eight conducted some form of peer evaluation during the 2013-14 SY. One of these districts used a peer assistance and review process only for the evaluation of new teachers and teachers who had previously been identified as having performance problems. In this district, peer evaluation ratings counted **for these teachers' 2013-14 performance evaluation**. One additional district trained representatives of the local teacher association to participate in observations with administrators. Two districts are planning to implement peer observation next SY.



### Evaluating Student Growth

The other key component of teacher evaluation under PERA is a measure of **teachers' contribution to student achievement growth**. As in most states, including measures of student growth is a new requirement for teacher evaluation in Illinois. For the 2013-14 SY, ISBE expected RTT districts (except CPS<sup>10</sup>) to pilot student growth measures for use in teacher evaluation. PERA regulations define student growth as “**a demonstrable change in a student's or group of students' knowledge or skills, as evidenced by gain and/or attainment on two or more assessments, between two points in time.**”<sup>11</sup> PEAC and ISBE created and promulgated guidance documents to help districts develop student growth measures for pilot use in evaluation.

The districts that provided information varied in how far along they were toward developing student growth measures and combining them with practice ratings to determine an overall summative evaluation rating. Twenty-four of the 28 districts had developed student growth measures for teachers for the 2013-14 SY. Of these districts, 11 included student growth measures along with practice ratings in an overall summative rating. However, only one district used the growth rating for high stakes (meaning counting the measure in the aggregate evaluation score) in the 2013-14 SY. Thirteen districts calculated student growth measures but did not combine them with practice ratings. Four districts did not provide sufficient information to ascertain their progress.

### *Assessments Used*

In order to calculate student growth, districts must determine what assessments they will use. PERA regulations state that districts must identify at least one Type I or Type II assessment and one Type III assessment for each teacher. If no Type I or Type II assessment is available for a teacher, districts may use two Type III assessments. Type I assessments measure a group of students in the same manner with the same potential assessment items, are scored by someone outside the district, and are administered either statewide or beyond Illinois. Type II assessments are those the district adopts and uses districtwide. Type III assessments must be rigorous, align with class curriculum, and be deemed by the evaluator and teacher to measure student learning.<sup>12</sup>

---

<sup>10</sup> PERA required CPS to include student growth in teacher evaluations used for administrative purposes for the 2013-14 SY.

<sup>11</sup> Illinois Administrative Code, Part 50, Sub A, Sec. 50.30:  
<http://www.isbe.net/rules/archive/pdfs/50ARK.pdf> Illinois

<sup>12</sup> <http://www.isbe.net/PEAC/pdf/student-growth-component-guidebook.pdf>

There appeared to be some confusion in several districts about what constitutes a Type I and Type II assessment. Several districts tended to view national assessments, such as the Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP)<sup>13</sup>, as Type II assessments, while others regarded them as Type I assessments. Many districts did not make a distinction between Type I and II assessments in developing growth measures. Of the 24 districts with growth measures for the 2013-14 SY, 22 used Types I and II assessments, and 2 used only Type III assessments. (One of these districts stated that its teacher contract did not allow the use of standardized assessments for evaluation.) Nine of these 22 districts used the state Illinois Standard Achievement Test<sup>14</sup> (ISAT) and Prairie State Achievement Examination<sup>15</sup> (PSAE)/Educational Planning and Assessment System (EPAS) assessments as Type I assessments. Districts used many different proprietary, externally developed and scored assessments. The most common were NWEA MAP, aimsweb,<sup>16</sup> Standardized Testing And Reporting,<sup>17</sup> Discovery,<sup>18</sup> iSystem to Enhance Education Performance<sup>19</sup> (iSTEEP), Study Island,<sup>20</sup> ThinkLink, and STI.<sup>21</sup> In addition, of these 22 districts, 15 used district-developed common assessments along with these externally developed tests; four districts used only externally developed and scored assessments as the Type I or II assessments, and one district used only district common assessments. Further, of these 22 districts, 8 districts gave teachers some degree of choice as to the assessments used. (For four districts it was unclear from the information provided whether teachers had choice or not.)

Seventeen of the 24 districts calculating growth used Type III assessments. The districts that were not using Type III assessments were typically starting by piloting student growth using Type I and II assessments and planned to begin using Type III assessments next year. Two districts used only Type III assessments. As might be expected, these assessments varied widely, ranging from textbook tests and teacher-made tests (the most common) to performance tasks, Individualized Education Plan (IEP) goals, writing rubrics, student portfolios and behavior checklists. Some districts also used screeners like Dynamic Indicators of Basic Early Literacy Skills<sup>22</sup> (DIBELS) as Type III assessments.

---

<sup>13</sup> <https://www.nwea.org/>

<sup>14</sup> <http://www.isbe.net/assessment/isat.htm>

<sup>15</sup> <http://www.isbe.net/assessment/psae.htm>

<sup>16</sup> <http://www.aimsweb.com/assessments>

<sup>17</sup> <http://www.starsamplequestions.org/>

<sup>18</sup> [http://www.discoveryeducation.com/pdf/assessment/Discovery\\_Education\\_Assessment\\_Research.pdf](http://www.discoveryeducation.com/pdf/assessment/Discovery_Education_Assessment_Research.pdf)

<sup>19</sup> <http://www.isteep.com/login.aspx>

<sup>20</sup> <http://www.studyisland.com/>

<sup>21</sup> <http://www.sti-k12.com/stias/>

<sup>22</sup> <https://dibels.org/dibels.html>

### *Measurement of Growth*

Only 2 of the 23 districts measuring growth based on Type I or II assessments used a statistically based (e.g., value-added) model to calculate student growth. All of the other districts used simple growth or a variation of Student Learning Objectives (SLOs) as their measurement model. (In many cases, the growth targets were set in a manner resembling SLOs, with teachers and principals agreeing on the targets of each teacher or grade/subject, even if the district did not call them SLOs.) One district set a single percent growth target for all teachers. Three districts provided insufficient information to tell what model they were using. One district had not yet determined its growth measurement model.

Only one of the districts using SLOs appeared to have conducted a central audit of the quality of the SLOs or their scoring. (Another district developed such a process and plans to use it next year.) Six of the districts using SLOs developed guidance documents or provided training to teachers and evaluators in setting and scoring SLOs. Two of these districts developed checklists or similar tools to help evaluators assess the quality of the SLOs proposed. Two districts planned to train next year. Three districts appeared to have no training or documentation that provides guidance on how to develop or score SLOs.

Few districts appeared to consider student characteristics in their growth measures. One of the districts that used a value-added model controlled for multiple student characteristics like low income, special education, and English language learner (ELL). The other district using such a model did not appear to take characteristics other than prior test scores into account. Five districts took student characteristics (primarily ELL and special education status) into consideration when setting growth goals or SLOs. Fifteen districts did not take student characteristics into account, though five of these planned to address this issue in the future. Three districts provided too little information to determine if they considered student factors in their student growth measures.

### *Data Quality*

Valid student growth measures start with using student test data that include as many of a **teacher's students as possible but do not include students for which the teacher is not responsible**. The districts participating in this portion of the study used various procedures to **ensure the proper students were counted for teachers' growth measures**. Overall, of the 24 districts measuring growth, 16 had some form of roster verification in which teachers and/or teachers and administrators ensured the correct students were **included for teachers' growth measures**. Five districts had formal roster verification processes in which teachers and administrators reviewed district-produced rosters. Districts using SLOs typically established the students to be included at the initial SLO meeting and at the end when the results were discussed. Two districts indicated problems with rosters that required manual review at the

end of the evaluation period. Five districts appeared to have not yet developed any process for roster verification, and three districts did not provide information about how they ensured valid student-teacher links.

However, while most districts addressed the issue of verifying students to be included in growth measures, fewer had considered other data quality issues. All districts required that students be present for both the pre- and post-test to be included, but relatively few had developed guidelines to address issues such as a minimum number of students needed to measure growth, student mobility, student absences, or team teaching. Several were developing or planned to develop such guidelines. Some of the smaller districts observed that given their relatively few teachers, they could handle issues on an ad hoc basis. Others reported that they would consider excluding students if the teacher provided a sound rationale for doing so. Except for the two districts using value-added measures for student growth, none of the other districts appeared to take potential test unreliability or student sample size and potential sampling error into consideration

#### *Overall Growth Rating*

Of the 24 districts that measured growth, 13 developed one overall growth score to combine with the practice rating to eventually produce an overall summative performance rating. Two districts did not combine each of their two growth scores, but kept them separate until averaged with the practice rating at the stage of determining a final overall summative rating. These districts split the weight given to growth between the two growth measures. Four of the districts that measured growth had not yet developed or finalized a method of combining the individual growth measures into an overall growth rating. Six districts did not provide sufficient information to determine whether they calculated an overall growth score or not.

Of the 13 districts that calculated a final growth rating, the most common method of combining growth measures was to convert the amount of growth on each measure into a rating on a one to four scale (e.g., less than target, on target, above target, much above target, or no growth, minimal growth, meets goal, exceeds goal) and then average these ratings and compare the average to a table specifying the overall growth rating, also on a one to four scale.<sup>23</sup> Other methods included rating scales based on the number of goals met or exceeded, the percentage of students showing growth, the percentage of students who met SLO objectives, or the average percentage growth of the students covered by the growth measures.

---

<sup>23</sup> For example, Excellent (4) = an average of 3.5 or higher, Proficient (3) = an average of 2.5 to 3.49, Needs Improvement (2) = an average of 1.5 to 2.49, and Unsatisfactory = an average below 1.5.

It was unclear in some cases how districts differentiated between levels of growth on the one to four scales they used. For example, few districts using rating labels such as “growth that was above target” and “growth was much above target” had quantitative guidelines to define these categories. Thus, it was difficult to determine how much higher student growth was for teachers rated “much above target” compared to teachers rated “above target.” In other cases, it appeared that teachers whose student growth exceeded goals by any amount were rated as “exceeds goal”; so, the potential exists for teachers who exceeded the goal by a small amount to receive the same rating as those who exceeded the goal by a large amount.

### Combining Practice and Student Growth Ratings

PERA will soon require that all districts produce a summative teacher evaluation score that combines practice and growth ratings. For the 2013-14 SY, 11 districts produced a summative evaluation score. To combine the practice and growth ratings, five of these districts used a decision matrix (similar to the one presented in the February 2013 PEAC document *Model Teacher Evaluation System—Creating a Summative Rating*<sup>24</sup>), while six used a weighted compensatory method (described in the PEAC document *Guidance on Creating a Summative Rating in Teacher Evaluation Systems*<sup>25</sup>). Of these 11 districts, 7 weighted practice at 70 percent and growth at 30 percent; 3 weighted practice at 75 percent and growth at 25 percent; and 1 did not provide this information. The 13 districts that did not combine practice and growth ratings planned to do so for the 2014-15 or 2015-16 SY. Most planned to use a weighted compensatory model and weight practice at 70 percent and growth at 30 percent.

### Communication With Teachers

In developing their new evaluation systems, districts recognized the importance of strategically communicating the new systems to their teachers. Most districts documented their teacher practice evaluation methods and communicated them to teachers via an evaluation handbook or similar document. Of the 28 districts that provided information, 13 shared documents that were quite comprehensive, including evaluation rubrics, timelines, examples of evidence, and explanations of how final practice ratings were to be calculated. An additional seven districts shared documents that were less comprehensive, but still provided substantial information about how the system worked. Many districts did not thoroughly document growth measures because the districts are still in the development phase. Four districts shared incomplete documents, and the remaining districts did not provide sufficient information to assess the quality of their documentation.

<sup>24</sup> <http://www.isbe.state.il.us/peac/pdf/guidance/13-9-te-model-summ-rating.pdf>

<sup>25</sup> <http://www.isbe.state.il.us/peac/pdf/guidance/13-11-te-summ-rating.pdf>

Almost all of the districts provided training on the new evaluation system to teachers. The scope and intensity of the training varied from one-on-one orientations with evaluators during a pre-conference to providing 40 hours of training on the rubrics and processes alongside the evaluators. Five districts reported developing web pages or intranet sites where teachers could access information about the evaluation process. Three districts were still in the process of establishing a communication plan and will be training teachers either over the summer or in the fall of 2014. Two districts provided no information about their training and communication efforts.

#### Summary of Teacher Evaluation System Progress and Potential Areas for Concern

This section summarizes the areas in which districts have made progress in implementing PERA-compliant teacher evaluation systems, including highlighting high-quality practices individual districts are using. The section concludes by identifying several areas of concern that state and district decision-makers may want to consider as districts continue to work on implementation.

#### *Areas of Progress*

1. Districts made progress on the professional practice component of their evaluation systems. All of the districts that provided information have established a practice evaluation component with the four levels required by PERA, adopted rubric-based rating systems aligned with state teaching standards, and required evaluators to participate in the state-sponsored training. In most districts, it appeared that the joint union-management committees charged with agreeing on evaluation practices worked productively to get this part of the evaluation process off the ground during the pilot years.
2. Almost all districts rated some or all teachers using the new professional practice evaluation processes for the 2013-14 SY.
3. Seventeen districts developed or were developing specialized versions of their rubrics for non-classroom certified staff, which have the potential to improve the validity of ratings of these staff compared to using a generic set of teacher rubrics.
4. **Many districts provided additional preparation to evaluators beyond the state's Growth Through Learning training.**
5. Almost all districts followed the minimum PERA requirements for the number of observations per year for both non-tenured and tenured teachers.

6. Most districts developed processes to measure or promote inter-rater agreement of teacher practice ratings.
7. The majority of districts developed a systematic method to combine practice and growth ratings to derive an overall summative evaluation rating, though only a minority did so for the 2013-14 SY (which is a pilot year for all but one district). All of those that provided information used or planned to use weights or methods consistent with PERA requirements or PEAC guidance documents.
8. Several districts developed innovative or especially high-quality teacher evaluation practices that might be useful for districts statewide. For example:
  - The Bloomington School district provides ongoing training to teacher evaluators at least two times a month.
  - The Calumet and Matteson School districts used co-observation (two evaluators observing the same teacher) to help promote inter-rater agreement.
  - CPS made available a wealth of material explaining the evaluation process and measures to its teachers. The district also developed model student performance tasks for most of its grades and subjects, which teachers use for their growth measure based on Type III assessments.
  - The Sandoval School district developed and provided to teachers and evaluators a detailed guidebook for student growth measures, with guidance on developing and evaluation student learning objectives.
  - The Unity Point School district developed guidelines for assessment timing, student attendance, and team teaching, and a process and tool for central office audit of SLO quality.
  - The Urbana School District uses the Danielson Framework for practice evaluation and has had Danielson Group trainers train and certify its evaluators. In addition to providing its educators with a detailed handbook, the district makes the TeachScape website, containing examples of teaching keyed to the Danielson Framework, available to teachers.

#### *Areas of Potential Concern*

1. Although most districts implemented measures to promote inter-rater agreement, a substantial number (at least 11) have not yet addressed this issue. Only two districts appear to have calculated a rater agreement index (i.e., the percentage of ratings that agree among different raters). This requires two evaluators to independently observe and rate a sample of teachers and thus provides information that districts can use to monitor the calibration of the raters and reassure teachers that their ratings are not dependent on the way a particular evaluator interprets the rubrics.
2. Most districts are following the minimum PERA requirements for the number of observations per year. However, they are likely conducting too few observations of tenured teachers to reliably assess teacher practice. Recent research<sup>26</sup> suggests that one observation by one observer has relatively low reliability.
3. Many districts are still developing methods of rating teachers based on student growth and combining these measures with the practice rating. As mentioned earlier, district progress has varied greatly, from a few still planning or developing measures to others that are ready to move forward with implementing a PERA-compliant system. (Note that one district has fully implemented a compliant system, as required.)
4. **Some districts' evaluation processes will use lagged growth data.** Nine districts used the state ISAT and PSEA/EPAS assessments as Type I assessments. Given that districts receive the results of these assessments long after the deadline for making evaluation ratings, growth on these assessments has to be measured using data from the prior two SYs. The growth measure thus reflects a different SY from the practice measure.
5. The methods of measuring student growth and combining the growth and practice measures vary widely. This raises a concern about the comparability of student growth measures and summative ratings across districts. Further, many districts appear to provide substantial discretion to teachers and evaluators in choosing assessments and defining growth goals. Some districts have not yet developed training or guidance in developing and scoring SLOs, and none appeared to have a process of auditing SLO quality by a central authority. These factors are likely to

---

<sup>26</sup> Rowan, B., Schilling, S.C., Spain, A., Bhandari, P., Berger, D., and Graves, J. (2013, December). *Promoting high quality teacher evaluations in Michigan: Lessons from a pilot of educator effectiveness tools*. Ann Arbor: University of Michigan Institute for Social Research; Bill & Melinda Gates Foundation (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET project's three-year study*. Seattle, WA: author.



result in some level of inconsistency in growth ratings across schools and potentially even within schools.

6. Data quality is an issue in some districts. Many of the smaller districts do not appear to have methods to ensure that teacher-student links are correct. In addition, most districts do not appear to be addressing issues such as a minimum number of students needed to measure growth, student mobility, student absences, or team teaching.
7. Few districts appear to be systematically accounting for the potential biases due to differences in the characteristics of students taught by different teachers, such as poverty or special education status. It is difficult for districts using growth targets or SLOs to do so, since no accepted methodology has yet been developed to precisely and consistently adjust goals or objectives for student characteristics associated with achievement.
8. Outside of the two districts using a value-added model, few if any districts have assessed the reliability and validity of their student growth measures. While one district using SLOs did examine the inter-rater agreement of SLO scoring, others appear to have not yet considered these issues. It should be noted that most of the districts are not likely to have staff with the expertise to address these issues and that there is not yet agreement in the field on how best to do so.

## **Principal Evaluation Systems**

### **Evaluating Leadership Practice**

For principals and assistant principals, PERA requires an evaluation of professional practice as one input to the overall summative performance rating. Similar as for teachers, evaluators use a rating scale or rubric to assess principals and assistant principals on professional practice. PERA regulations require these to align with the Illinois Performance Standards for School Leaders.

#### *Practice Rubrics, Observations, and Additional Evidence Districts Used*

Sixteen of the 28 RTT districts that provided information, and each of the other 4 districts, reported using the Illinois Performance Standards for School Leaders rubric<sup>27</sup> without modification. Four districts used adapted or modified versions of this rubric. Five used other rubrics (primarily the Governors State University model), and one used a general rubric applying to all district administrators.

Almost all districts reported making two formal observations of principals. Two districts did not share information about the frequency of principal observations. The observations typically lasted from 45 minutes to 1 hour, though several districts reported that the observations would last as long as the particular activity (e.g., a conference with a teacher or staff meeting). Most districts also made use of an indefinite number of informal observations, though a few (two) of these did not use the informal observations for evaluation purposes. Four districts included walk-throughs as informal observations.

For a more comprehensive picture of principal practice, all districts required or allowed principals to provide artifacts relevant to the leadership standards, such as meeting agendas or minutes, professional development plans or records, communications with parents or teachers, school improvement plans, and various additional data reports. However, most districts did not provide the research team with documentation that specified the exact artifacts principals should submit. Five districts required or encouraged principals to prepare a portfolio of relevant artifacts.

Four districts using the Governors State University model included the Vanderbilt Assessment of Leadership in Education<sup>28</sup> (Val-Ed), a 360 assessment, as part of the practice component. In this model, the Val-Ed ratings counted for 30 percent of the final summative rating. One other district offered principals the option of using results of a school climate/culture survey as evidence about leadership practice.

#### *Evaluator Training*

Noted in the teacher section, training of evaluators is critical to ensure quality practice ratings. Trained raters are more likely to apply the practice rubrics accurately and consistently. Twenty-seven of the 28 RTT and all 4 of the other districts reported that all those who evaluated principals took and passed all of the required training provided by state-

---

<sup>27</sup> <http://illinoisasa.wikispaces.com/file/view/IL+Principal+Evaluation+State+Model+Summary+with+Appendices+21512.4.pdf>

<sup>28</sup> <http://www.valed.com/>

sponsored Growth Through Learning training. Twelve of the total 32 districts provided additional training beyond that sponsored by the state.

State guidelines require additional, rubric-specific training for principals, assistant principals, and other evaluators when a district uses a rubric other than the state model. Of the six districts that used rubrics that differed substantially from the state model, five provided additional training to evaluators on the rubric, and one did not provide information on additional training.

### *Inter-Rater Agreement*

Although assessing inter-rater agreement is important for accuracy of ratings and credibility among principals, no districts appear to have assessed the reliability of principal practice ratings by calculating inter-rater agreement statistics. Eight of 32 districts did take actions to promote agreement, such as having regular calibration meetings or having periodic discussions among evaluators. The remaining districts either did not take such steps or did not provide the study team with the information. However, six of these districts only had one evaluator, making it difficult to check on inter-rater agreement.

### *Overall Rating of Leadership Practice*

After evaluators used their judgment to rate each of the practice standards or components, most (27) districts used a defined, mechanical process to **determine the principal's aggregate leadership practice rating**. Similar as for teachers, districts most commonly (1) summed or averaged the one to four ratings on each component, then compared the result to a table that provided score ranges required for each on the four overall practice levels, or (2) required that a certain number of components be rated at each level (e.g., to be rated at the highest level, the principal needs 13 components rated at the highest level and no more than 1 at the needs improvement or **unsatisfactory level**), **which parallels the state's recommendation**. For the remaining districts, three allowed the evaluator to use his or her judgment, and three did not provide the study team with sufficient information.

Unlike the teacher aggregate practice rating, the state requires districts to include a principal self-assessment as part of the aggregate principal rating. However, some (10 RTT and 3 other districts) did not appear to use the self-assessments in determining the aggregate principal practice or final rating. Nine districts used the self-evaluation as an input in calculating practice standard or component ratings, without specifying its weight, while seven districts used the self-evaluation as an explicit input into the final overall (practice, growth, and additional factors) rating—weighting it at 10 percent. Four districts did not provide sufficient information to determine their use (or lack thereof) of the self-assessment.

#### *Differences Between Principal and Assistant Principal Practice Evaluation*

Districts assess principal and assistant principal practice skills very similarly. With one exception, all other districts used substantially the same rubrics and process to evaluate assistant principals as they used for principals. Note that for districts using the Governors State University model, there was potentially more opportunity to differentiate, since the model allows for setting and evaluating different practice goals for each administrator. One major difference between principal and assistance principal practice evaluations was that in most districts, principals evaluate the assistant principals. However, in five smaller districts, the superintendent, assistant superintendent, or another district-level administrator observed or evaluated the assistant principals.

#### *Evaluating Student Growth*

For the 2012-13 SY, the state required that Illinois districts implement an evaluation process for principals and assistant principals that included a student growth measure. Thus, during the 2013-14 SY, districts should have been in their second year of implementing the system. As they did for teacher growth measures, districts should measure principal growth measures by assessing the change in student achievement between two points in time. State guidelines allow districts to use additional outcome measures such as ISAT attainment, graduation rates, or increases in attendance to evaluate student growth.<sup>29</sup>

Thirty districts reported using student growth as a part of their principal evaluation systems and combined student growth measures with practice and other measures to determine an overall aggregate evaluation rating for the 2013-14 SY. Two districts did not appear to be using student growth.

#### *Assessments Used*

The state requires districts to evaluate principal student growth ratings by using any assessments that meet the definition of Type I or Type II assessment (described earlier). If **these assessments do not cover most students in a principal's school, the state allows the district to use Type III assessments** (noted above). Most districts (28) used Type I and/or II assessments. Based on provided data, the **team was unable to discern three districts'** assessments. The remaining district was in the process of reconsidering what assessments it would use for the 2013-14 SY; however, for the previous SY, the district used Type I and II assessments.

---

<sup>29</sup> [http://www.isbe.state.il.us/peac/word/peac\\_prin\\_eval\\_model.pdf](http://www.isbe.state.il.us/peac/word/peac_prin_eval_model.pdf)

Similar to the teacher evaluation, many districts appeared confused in the distinction between Type I and Type II assessments. Several districts tended to view national assessments, such as the NWEA MAP,<sup>30</sup> as Type II assessments, while others regarded them as Type I assessments. Many districts did not make a distinction between Type I and II assessments in developing growth measures. More specifically, 12 of the 32 districts used the state ISAT or PSEA/EPAS assessments as Type I assessments. Using one-year lagged assessment data, the districts defined student growth as an increase in the percentage proficient score level between two years. In addition, many districts used externally developed and scored assessments, such as NWEA MAP, aimsweb, and Discovery. Eleven districts also used district-developed common assessments along with externally developed assessments. However, 13 districts reported they did not use such assessments; 6 did not provide enough information to determine; and 1 district had not yet decided on assessments. Further, four districts reported using Type III assessments, mostly to measure growth for special education or English learner student populations. One district reported determining student growth for principals by rolling up teacher-level SLO results (using Type III assessments) to the principal level.

### *Measurement of Growth*

Only 3 of the 32 districts used a statistically based growth model to calculate student growth measures based on Type I or II assessments. The remaining districts set growth targets or goals based on simple growth calculations (i.e., difference in scores on pre- and post-tests). Two of the three districts using statistical models also set such targets. In 14 of the districts, principals had some degree of choice over both the assessments used and the growth targets. In these districts, the targets resembled school-level SLOs. In six districts, principals influenced either assessment selection or goals set, but not both. Only four districts used a uniform set of assessments and goals for all principals. The study team was unable to make this determination for eight districts due to limited data.

Districts varied in the standards they used to rate student growth. For example, while several districts used a 4-point rating scale where “3” equals “meets target,” others used a 4-point rating scale where “2” equals “on target.” Note that these two rating scales do not align. In addition, few districts seemed to have rubrics or other documents to guide evaluators as they rated growth. Further, few districts appeared to have an objective method of defining the ratings given to each measure, goal, or target.

---

<sup>30</sup> <https://www.nwea.org/>

Few districts appeared to take student characteristics (i.e., low income, IEP, ELL status) into account in their growth measures. Four districts noted that they considered characteristics, such as ELL and special education status, when setting goals or targets, but it was not systematic. However, two districts reported that they can or do set separate goals to focus on specific groups of students, while three districts were considering ways of accounting for student characteristics in the 2014-15 SY.

#### *Data Quality*

Discussed earlier, **data quality is important in using student growth measures in educators' evaluation ratings.** Relatively few (six) districts had procedures to verify school rosters to ensure the proper students were **counted for in principals' growth measures.** Other than requiring students to be in the school between the dates of pre- and post-tests, only four districts appeared to have systematically addressed issues like student absences and movement between schools. In addition, it appeared that few districts have addressed the potential issues related to measurement error in growth scores.<sup>31</sup>

#### *Overall Growth Rating (for student growth only)*

To combine growth measures, most districts converted the amount of growth on each measure into a rating on a scale of one to four (e.g., less than target, on target, above target, much above target; or no growth, minimal growth, meets growth, exceeds growth), then averaged the results and compared the average to a table specifying the overall growth rating, also on a scale of one to four scale. Other methods included basing the overall growth rating on the number or proportion of goals met or exceeded (e.g., meets no targets /negative growth on one or more measures = no growth; meets one or two targets and has negative growth on no more than one measure = minimal growth; meets or exceeds target for majority of measures with no negative growth = meets growth; reaches or exceeds the target for majority of measures or meets all targets = exceeds growth) and basing the overall growth rating on the average percentage growth or the percentage of students meeting targets. Eight districts did not provide sufficient information to determine how they combined measures of growth into an overall student growth rating.

---

<sup>31</sup> However, because of the individualized nature of the goal/target setting process and the reliance on judgment in rating growth, it would be difficult for most districts to take potential measurement or student sampling error into consideration in assessing the quality of the student growth measures applied to principals and assistant principals.

### *Other Measures*

Thirteen of the 32 districts also considered other student outcome information as part of the student growth component. However, 14 did not consider additional outcome measures, while 5 districts did not provide sufficient information to determine. In addition, eight districts also included non-student achievement measures such as walk-through ratings, school climate survey results, staff absences, parent satisfaction survey results, quality of teacher evaluation (completed by the principal) in principal evaluation. Note that it is unclear what percentage of these districts were using the non-student achievement measures as part of the growth score. Some districts may have used these measures as part of the practice score or as an independent factor (contributing to the final aggregate evaluation score).

### *Differences Between Principal and Assistant Principal Growth Evaluation*

Although PERA guidance documents suggest that districts may want to apply different growth measures, outcomes, or weights to assistant principal growth evaluation scores, only half of the districts appeared to take advantage of this flexibility. Of those that did, the most common modification was to give higher weights to measures like out-of-school suspensions or student attendance.

### *Combining Practice and Student Growth Ratings*

Most (29) districts used a weighted compensatory method to combine growth (including other student outcomes) and practice ratings for principals. Of these districts:

- ▶ Eleven weighted the practice rating as 75 percent and the growth rating as 25 percent.
- ▶ Four weighted the practice rating as 70 percent and the growth rating as 30 percent.
- ▶ Seven weighted the practice rating between 60-65 percent, the growth rating between 25-30 percent, and the self-evaluation as 10 percent.
- ▶ Six used different weighting schemes, ranging from weighting the practice rating as 50 percent and the growth rating as 50 percent to weighting the practice rating as 70 percent, the growth rating as 25 percent, and other outcomes as 5 percent.

In addition, one district used the PEAC-developed state summative rating matrix, while another district allowed principals to choose between using this matrix and a weighted compensatory method. Further, one district appeared to combine the components judgmentally, while another district did not provide enough information to determine.

#### Communication With Principals

All districts appeared to have informed principals of PERA performance expectations and criteria for evaluation. However, the amount of information provided varied. While a few districts developed a principal evaluation guidebook that described both the practice and growth components in detail, most provided less information, especially regarding growth measurement. This may have been because the growth goals were set individually with principals in initial goal-setting meetings.

#### Summary of Principal Evaluation System Progress and Potential Areas of Concern

This section summarizes district progress in implementing PERA-compliant principal evaluation systems, including highlighting best practices individual districts are using. The section also identifies some areas of concern that state and district decisionmakers may want to consider as districts continue to implement and refine their systems.

#### *Areas of Progress*

- ▶ The majority of districts are using the Illinois Performance Standards for School Leaders rubric, which align to the Illinois Professional Standards for School Leaders. Most districts using another rubric provide training on their specific rubric.
- ▶ Almost all districts appeared to have informed principals and assistant principals about the PERA evaluation standards and processes.
- ▶ Most districts developed procedures and are using student growth as part of their principal evaluation systems.
- ▶ The majority of districts developed a systematic method for combining practice and growth ratings to derive an aggregate evaluation rating and are using weights and methods compliant with PERA.
- ▶ Several districts developed innovative or especially high-quality teacher evaluation practices that might be useful to share with districts statewide. For example:
  - Brooklyn, a small district with one person acting as both superintendent and principal, used a superintendent from a neighboring district to **evaluate its principal's professional practice**.
  - CPS developed extensive web-based resources and FAQs for its principals that explained the principal evaluation process and measures clearly and succinctly.



- West Richland explicitly **included the quality of school leaders' evaluation** of teachers in its principal evaluation system.
- Zion provided rubrics for observations of principals conducting a teacher evaluation conference and a professional development meeting.

### *Areas of Potential Concern*

- ▶ Most districts did not provide as much documentation on their principal evaluation systems compared to their teacher systems. Further, districts lacked thorough documentation on growth measurement.
- ▶ While some districts have taken actions to promote inter-rater agreement, many have not. Although inter-rater agreement is harder to define and measure for principals compared to teachers, districts with more than one principal evaluator should be making an effort to ensure that evaluators of principals are interpreting the rubrics in a similar way and using similar evidence to assess practice.
- ▶ Few districts appeared to assess the potential reliability and validity of their student growth measures.
- ▶ Districts varied substantially in the standards used to rate growth, and many lacked rubrics objectively defining the amount of growth to distinguish rating levels. This suggests that growth ratings may not be comparable from district to district. Since most districts allowed principals and evaluators to choose assessments and set individualized goals, it is likely that ratings may not be comparable from school to school within those districts.
- ▶ Many districts have not systematically addressed data quality issues, such as a minimum number of students (need to measure growth), student mobility, and student absences.
- ▶ A few districts did not use an objective, uniform method to calculate the practice rating, leaving the method of combination to evaluator judgment. This not only allows for differential treatment across principals, but also makes it difficult for principals to understand how the evaluator determined the practice rating

## **Concluding Remarks**

Implementing new evaluation systems can be a major undertaking for school districts, particularly small districts. Many Illinois districts are relatively small and thus often face capacity challenges as they roll out new evaluation systems. One of the largest challenges districts faced was the implementation of growth measures. The additional pilot year will be essential for these districts to work to improve the quality and coverage of the measures in order to reliably use these metrics in their evaluation systems.

Notably, state guidance documents and evaluator training appeared useful to many districts as they began to implement the new evaluation systems. However, additional dissemination of resources to districts may prove beneficial. Further, technical assistance from the state, regional education agencies, and universities will likely be needed to enable all of the districts to comply fully with PERA for the 2016-17 SY.

Since PERA regulations and state guidance allowed districts a great deal of flexibility in developing the educator evaluation systems, districts were able to develop evaluation practices to meet local needs, including stakeholder acceptance (in some instances). While this flexibility allowed districts to design systems to better meet their individual needs, it is difficult to compare evaluation systems (and particularly the growth component) across districts.

# Survey Data

This chapter examines teacher, principal and assistant principal, and principal evaluator perceptions of the PERA-compliant educator evaluation systems. To collect this information, the research team surveyed the educators and staff in spring 2014. Team members developed the survey instruments based on a review of PERA guidance documents, existing surveys **developed by the CCSR, and from studies of other states' evaluation systems, as well as the** teacher reactions to performance evaluation scales developed by Heneman and Milanowski.<sup>32</sup> These survey data provide insight into teacher, principal/assistant principal, and evaluator **perceptions of the evaluation processes and can inform ISBE's assessment of districts'** progress toward implementing PERA-compliant evaluation systems.

To administer the teacher survey, the research team worked with CCSR and the University of Chicago Impact program to include additional items about the new evaluation process on the 5 Essentials teacher survey. In the spring of 2014, the CSSR administered the survey to teachers in CPS, and Impact administered the survey to teachers in the rest of the state. Data analysis used 44 items from the survey. The overall response rates for the 5 Essentials survey was 81 percent for CPS and 58 percent for the rest of the state. A total of 22,336 teachers from the 35 study districts, which comprised approximately 47 percent of our total sample, provided responses.

**In addition to the teacher survey, the research team designed surveys to collect principals', assistant principals', and principal evaluators' perceptions of the evaluation processes in study districts.** To administer these web-based surveys, the team requested email addresses for principals, assistant principals, principal evaluators, and teacher evaluators from each study district. Using the email addresses received from the districts, the team randomly administered a web-based survey to approximately half of the principals/assistant principals about their perceptions of their own evaluation process. Of the 131 principals and assistant principals invited to participate in this survey, the team received 100 responses from 22 districts for a response rate of 76 percent. Fifty-eight percent of the respondents were principals, and 42 percent were associate or assistant principals. The team administered a separate web-based survey about perceptions of the teacher evaluation process to the

---

<sup>32</sup> Heneman, H.G. III, and Milanowski, A.T. (2003). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education*, 17(2), 173-195.

remaining principals and assistant principals in the sample. Of the 118 principals and assistant principals invited to participate in this survey, 88 responded, representing 19 districts, for a response rate of 75 percent. Sixty-one percent of the respondents were principals, and 39 percent were associate or assistant principals. In addition, to reduce burden on potential respondents in Chicago, the team asked the CCSR to include questions on **principals' evaluation experiences on a survey it was already conducting in CPS. Of the** surveys administered, 287 principals, 297 assistant principals, and 131 school leaders (who did not provide a job title) responded to the survey. The team also administered a survey to principal evaluators (e.g., district superintendents) to assess evaluator perceptions of the principal evaluation process. The team received responses from 64 evaluators in 25 districts. The respondents comprised 20 superintendents; 20 deputy, associate, or assistant superintendents; 15 network chiefs or deputy chiefs in CPS; and 9 other district administrators. Though the number of evaluators receiving the survey was small relative to teachers, principals, and assistant principals, 73 percent of principal evaluators for whom the team had contact information responded to the survey.

Since the study districts varied widely in size—ranging from a single school to over 500 schools in the largest district—there was substantial variation in the number of respondents from each district in the sample. Had the team reported survey results at the respondent level, the largest district would have dominated the results. This would make it difficult to understand how other districts perceived the evaluation systems. Therefore, in collaboration with ISBE, the team decided to make the unit of analysis the district and report results in terms of district averages or percentages of respondents agreeing with particular survey items. This approach gives each district an equal weight no matter the number of respondents. The limitation of this approach is that the results do not show what proportions of teachers, principals, assistant principals, and evaluators across all study districts held various opinions about their evaluation systems.

The results support the following general conclusions:

1. Evidence from the surveys indicates that in most districts, a majority of teachers, principals, and evaluators had a strong level of understanding of the professional practice rubrics, but not necessarily of the student growth measure or how the measures are combined.
2. Teacher and principal responses by district indicated that on average they perceived the professional practice evaluation as fair, but had mixed perceptions of the fairness of the growth measures.
3. In most districts, a majority of teachers, principals, and evaluators perceived that the evaluations influenced either teacher or principal professional development.
4. In most districts, a majority of teachers, principals, and evaluators indicated that either teachers or principals changed their instructional or leadership practices as a result of the evaluation process, but principal evaluators were more optimistic about **changes in principals' practices than teacher evaluators were about changes in teachers' practices. In most districts, teachers, principals/assistant principals, and principal evaluators were also optimistic about the future impact on practice and student learning, but had concerns about whether the results were worth the extra effort.**

The remaining chapter discusses the detailed findings based on a district-level analysis of **teachers', principals and assistant principals', and evaluators'** survey responses. It presents, in the following order, the teacher, teacher evaluator, principals/assistant principal, and principal evaluator survey results. The chapter concludes with a discussion of some additional patterns in the results.

## Teacher Perceptions of the Evaluation Process

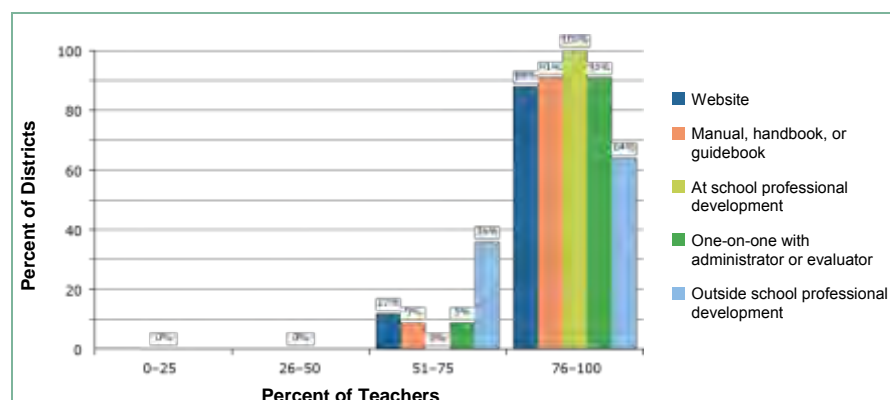
### Methods of Providing Teachers With Information on Evaluation Process

Communication about the evaluation process is an important prerequisite for educators to understand the systems and the performance expectations they represent. To this end, the research team asked teachers about the communication materials districts used to inform them about the evaluation systems and the perceived usefulness of the materials.

- ▶ In the majority of districts, teachers reported receiving multiple forms of information on the teacher evaluation system. It appeared **that districts' primary** means of disseminating information on the evaluation system was through a website or an evaluation manual, handbook, or guidebook. (See Figure 1 below.)
  - In every district, more than 51 percent of teachers reported receiving information from at least one source on the evaluation system.
    1. In 88 percent of the districts, 76 percent or more of teachers reported receiving information on the evaluation system through a website.
    2. In 91 percent of the districts, 76 percent or more of teachers reported receiving information on the evaluation system in a manual, handbook, or guidebook.
    3. In 91 percent of the districts, 76 percent or more of teachers reported receiving information on the evaluation system through one-on-one interaction with their administrator or evaluator.
  - In all districts, 51 percent or more of teachers reported receiving information on the evaluation system through professional development sessions at the school or outside it.

In every district, 76 percent or more of teachers reported receiving information on the evaluation system through in-school professional development, while in only 64 percent of districts did 76 percent or more of teachers report receiving information on the evaluation system through professional development outside the school.

**Figure 1.** Percent of Teachers Within Districts Receiving Information on Evaluation Systems<sup>33</sup>



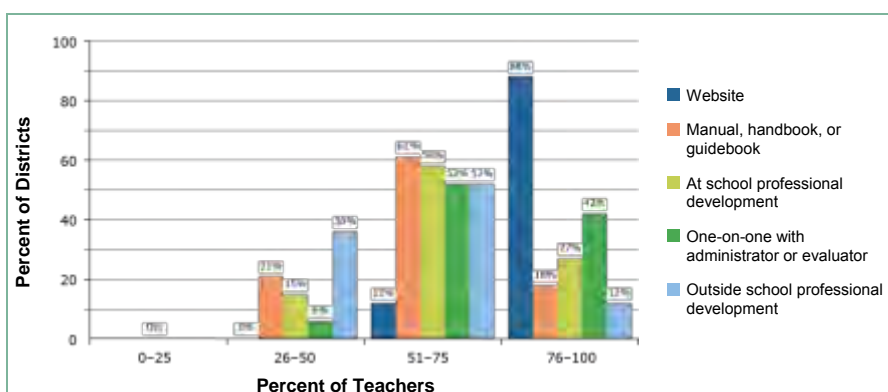
### *Usefulness of Various Communication Methods*

- ▶ Though the majority of teachers within each district reported receiving resources on the evaluation system, it appeared that the degree to which they found the information “useful” or “very useful” varied across information sources. (See Figure 2 below.)
- ▶ The resources perceived as useful in the most districts were websites and one-on-one meetings with administrators or evaluators.

<sup>33</sup> Reading the figures: In most of the figures, we show the percent of districts in which 0-25, 26-50, 51-75, and 76-100% of the respondents answered survey questions in various ways (e.g., reported receiving information, amount of time spent, or agreeing or strongly agreeing to statements about the evaluation process). For example, Figure 1 shows that in 12% of the districts with teachers who responded to the survey, between 51% and 75% of teachers reported receiving information on evaluation systems through a website. In 9% of these districts, between 51% and 75% of teachers reported receiving information on evaluation systems through a manual, handbook, or guidebook. In 0% of these districts, between 51% and 75% of teachers reported receiving information on evaluation systems through professional development sessions at their school. In 9% of these districts, between 51% and 75% of teachers reported receiving information on evaluation systems through one-on-one meetings with an administrator or evaluator. In 36% of these districts, between 51% and 75% of teachers reported receiving information on evaluation systems through professional development sessions outside the school. The percentages of districts in which 76% to 100% of teachers reported receiving information on each source is shown in the next set of bars. These figures are intended to portray the degree to which there is consensus across districts and respondents within districts about aspects of the evaluation systems. In most cases, where more districts have 51-75% and especially 76-100% of respondents agreeing, this indicates that evaluations systems are being perceived positively overall in the study districts.

- In every district, 51 percent or more of teachers reported that the website was useful or very useful, and in approximately 89 percent of the districts, at least 76 percent of teachers reported the same.
  - In 94 percent of districts, 51 percent or more of teachers reported that one-on-one interaction with their evaluator or school administrator was useful or very useful.
- In a smaller number of districts, most teachers perceived the other resources as useful or very useful.
- In 79 percent of the districts, 51 percent or more of teachers reported that the evaluation manual, handbook, or guidebook was useful or very useful.
  - In 85 percent of districts, at least 51 percent of teachers reported that the in-school professional development on the evaluation process was useful or very useful.
  - In 64 percent of districts, 51 or more percent of teachers reported that professional development outside the school was useful or very useful.

**Figure 2.** Percentages of Teachers Within Districts Rating Information Sources as Useful



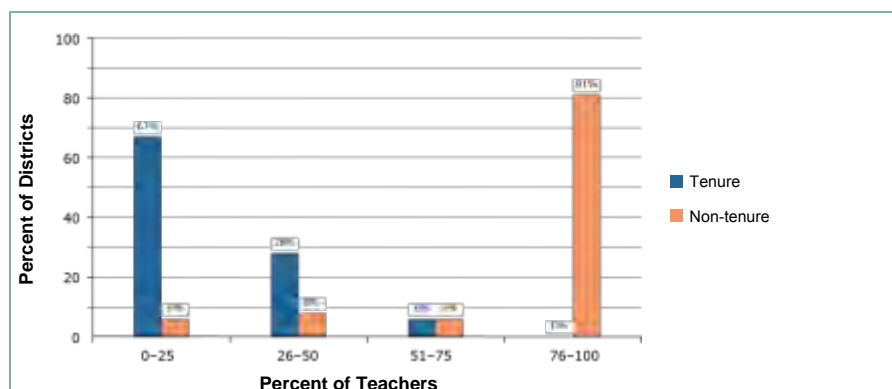


### Frequency of Teacher Observations

PERA requires that most non-tenured teachers receive at least two formal observations each SY and that proficient or above tenured teachers receive two observations every other year. To assess how well districts were fulfilling these requirements, teachers responded to a question about the number of observations they received in SY 2013-14. (See Figure 3 below.)

- ▶ In 80 percent of the districts with responding teachers, 76 percent or more of the non-tenured teachers reported receiving two formal observations. In the average district, 83 percent of these teachers reported receiving two observations.
- ▶ Because most tenured teachers would be on a two-year evaluation cycle, the study team did not expect most of them to report having received two formal observations. In the average district, 22 percent of these tenured teachers reported receiving two observations. In 67 percent of districts, less than 25 percent of tenured teachers reported receiving two or more observations, which is less than expected if about half of the tenured teachers received evaluations during the 2013-14 SY. It is possible that these districts piloted new evaluation systems mostly for non-tenured teachers.

**Figure 3.** Percentage of Teachers Within Districts Reporting Receiving Two or More Observations<sup>34</sup>



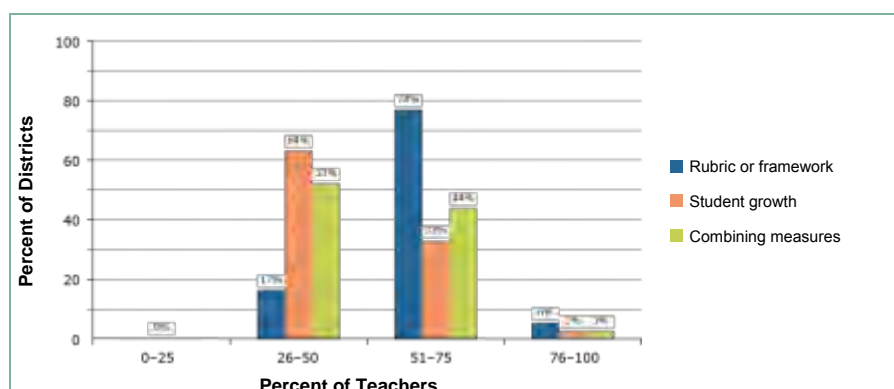
<sup>34</sup> Figure 3 reads: In 67% of districts, between 0% and 25% of tenured teachers reported receiving two or more observations, and in 6% between 0% and 25% of non-tenured teachers reported receiving two or more observations. In 28% of districts, between 26% and 50% of tenured teachers reported receiving two or more observations, and in 8% between 26% and 50% of non-tenured teachers reported receiving two or more observations. In 6% of districts, between 51% and 75% of tenured teachers reported receiving two or more observations, and in 6% between 51% and 75% of non-tenured teachers reported receiving two or more observations. In no districts did 76% or more tenured teachers report receiving two or more observations, and in 81%, 76-100% of non-tenured teachers reported receiving two or more observations.

### Teacher Understanding of the Evaluation Components

Teachers' understanding of their evaluation system is a good indicator of the quality of district communication efforts and likely affects the acceptance, success, and sustainability of the system. To gauge **teachers' understanding** of the evaluation systems, the team examined teacher responses to questions asking about their understanding of the rubric or framework, the student growth measure, and the method for combining the two measures into a final rating. **Examining teachers' understanding of the evaluation process yielded insights into** which evaluation components districts should focus their communication efforts.

- ▶ In almost all of the districts, most teachers rated communication methods as useful; however, in fewer districts, most teachers reported that they had a strong understanding of the components of the evaluation process, especially the student growth measure. (See Figure 4 below.)
- ▶ In general, it appeared **teachers' understanding of the student growth component and the districts' process for combining measures** was not as strong as their understanding of the evaluation framework.
  - In 83 percent of the districts, 51 percent or more of the teachers reported a strong or very strong understanding of the evaluation rubric or framework. However, in only 6 percent of districts, 76 percent or more of teachers reported a strong or very strong understanding of the evaluation rubric or framework.
  - Conversely, in only 35 percent of districts did 51 percent or more of teachers report a strong or very strong understanding of their student growth measure. And, in only 3 percent of districts did 76 percent or more of teachers report a strong or very strong understanding of the student growth measures.
  - In 47 percent of districts, 51 percent or more of teachers reported having a strong or very strong understanding of how their districts combine measures to create a final evaluation rating. However, in only 2 percent of districts did 76 percent or more of teachers report a strong or very strong **understanding of their district's process**.

**Figure 4.** Percentage of Teachers in Districts Rating Their Understanding of System Components as Strong or Very Strong

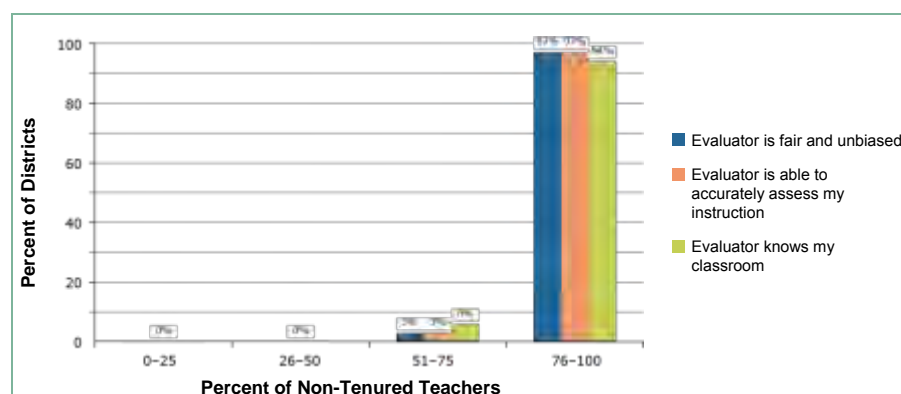


#### Teacher Perceptions Related to Credibility of Their Evaluator and Fairness of the Rubric

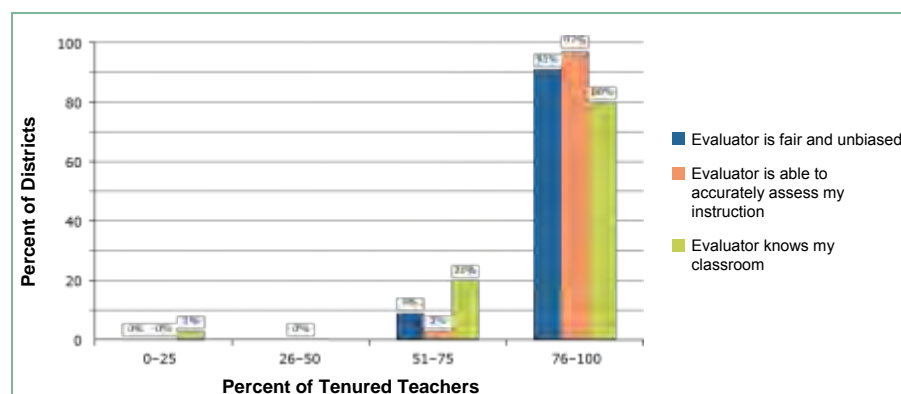
Understanding teacher perceptions of the credibility of their evaluator and rater consistency across teachers is important for assessing whether teachers being evaluated perceive the process as fair. It is also an indicator of how likely teachers are to accept feedback from the evaluator and use it to improve practice. Low evaluator credibility can undermine the intended outcome of implementing high-stake evaluation systems (i.e., improved educator practice and ultimately student achievement). This study asked teachers about their perceptions of the credibility of their evaluator, the consistency of evaluator ratings across teachers, and the overall fairness of the professional practice rubric and growth measures.

- The majority of tenured and non-tenured teachers reported that their evaluator was fair and unbiased, understood their classroom, and was able to accurately assess their performance. Non-tenured teachers were slightly more likely to agree with these three statements than tenured teachers. (See Figures 5a and 5b.)

**Figure 5A.** Percentage of Non-Tenured Teachers in Districts Rating Evaluators as Fair and Credible



**Figure 5B.** Percentage of Tenured Teachers in Districts Rating Evaluators as Fair and Credible



- Teachers also tended to agree that the professional practice rubric was a fair representation of good teaching. In all districts, 51 percent of more teachers agreed or strongly agreed. In the average district, 79 percent of the teachers agreed or strongly agreed.

### *Teacher Perceptions of Evaluator Consistency Across Teachers*

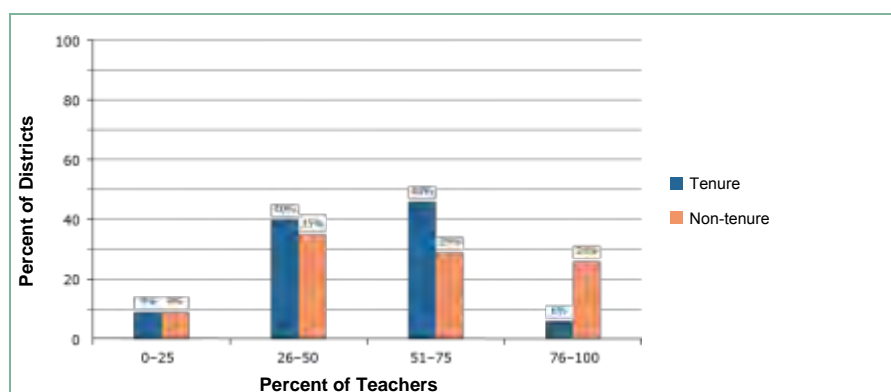
- ▶ Somewhat unexpectedly, even though in the vast majority of districts most teachers perceived their evaluators as knowledgeable and credible and the rubric fair, both non-tenured and tenured teachers tended to agree that the ratings they received depended more on the evaluator than their practice.
  - In 91 percent of districts, 51 percent or more of tenured teachers agreed or strongly agreed that their ratings depended more on the evaluator than on teaching practices.
  - In 65 percent of districts, 51 percent or more of non-tenured teachers agreed or strongly agreed that the ratings depended more on the evaluator than their practice.

### *Teacher Perceptions of the Fairness of the Growth Measures*

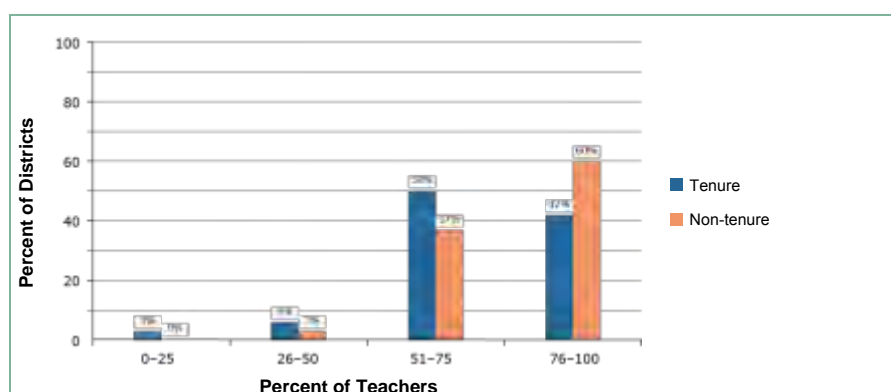
- ▶ **In comparison to teachers' perceptions of their evaluators** and the rubric, teachers were less inclined to agree that the growth measure based on Type I or II assessments is fair. (See Figure 6a.)
  - In 51 percent of districts, 51 percent or more of tenured teachers reported that they agreed or strongly agreed with the statement “**the measures of student growth [based on Type I or II assessments] are a fair representation of my students' learning.**” Similarly, in 56 percent of districts, 51 percent or more of non-tenured teachers agreed or strongly agreed with the same statement.
- ▶ Teachers were more likely to perceive the growth measure based on Type III assessments as fair. (See Figure 6b.)
  - In 97 percent of the districts, 51 percent of the non-tenured teachers agreed or strongly agreed that the student growth measures based on the **Type III assessments were a fair representation of students' learning.** In 90 percent of the districts, 51 percent of the tenured teachers agreed or strongly agreed.
- ▶ In addition, there is a slight split in opinion on student growth measures between tenured and non-tenured teachers. Generally, tenured teachers by district were less inclined than non-tenured teachers to perceive the growth measure as fair.
  - In the average district, 59 percent of the non-tenured teachers agreed or strongly agreed that the student growth measured based on Type I and II assessments are a fair representation of their students learning, compared

to 49 percent of the tenured teachers. For the growth measure based on Type III assessments, in the average district, 78 percent of non-tenured and 69 percent of tenured teachers agreed or strongly agreed.

**Figure 6A.** Percent of Teachers Agreeing That Measures of Student Achievement Growth Based on Type I/II Assessments Are a Fair Assessment of Their Students' Learning



**Figure 6B.** Percent of Teachers Agreeing That Measures of Student Achievement Growth Based on Type III Assessments/Performance Tasks Are a Fair Assessment of Their Students' Learning

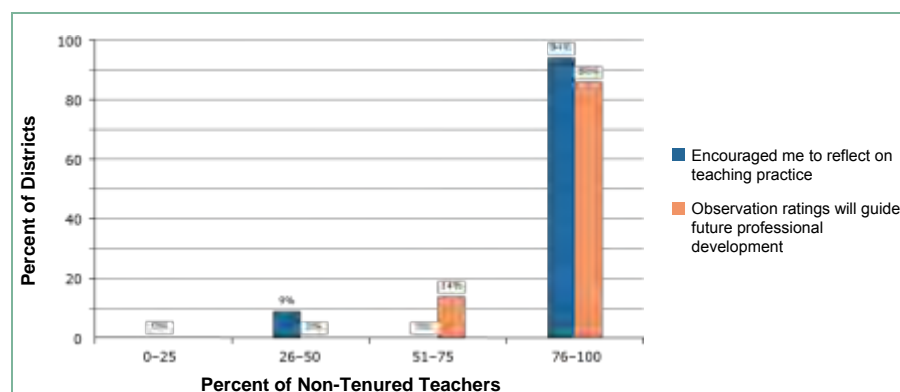


### Teacher Perceptions of Evaluation Impact on Professional Development

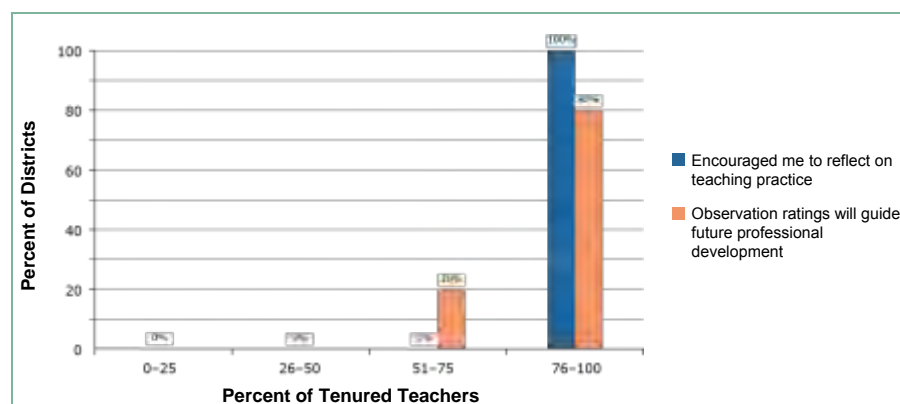
A key goal of teacher evaluation systems is to assess and improve teacher practice and student learning. One mechanism for improving teacher practice is to provide targeted feedback to teachers to guide professional development. To address whether the evaluation system informs teacher professional development, the survey asked about the feedback teachers received and how it might influence their professional development plans.

- ▶ In the majority of districts, both non-tenured and tenured teachers reported that post-observation feedback would guide future professional development. (See Figures 7a - d below.)
  - In 88 percent of districts, 76 percent or more of teachers reported they agreed or strongly agreed that the post-observation feedback identified specific areas for improvement, included guidance or suggestions on making improvements to instruction, and will ultimately help improve their instruction.
  - In 94 percent of districts, 51 percent or more of non-tenured teachers reported that they agreed or strongly agreed that the observation process encouraged them to reflect on their teaching practice. The responses among tenured teachers were even more positive. In 100 percent of districts, 76 percent or more of tenured teachers reported they agreed or strongly agreed with the same statement.
  - In all districts, 51 percent or more of non-tenured teachers reported they agreed or strongly agreed that their observation ratings will guide future professional development. Similarly, in all districts, 51 percent or more of tenured teachers reported they agreed or strongly agreed with the same statement.

**Figure 7A.** Percent of Non-Tenured Teachers Reporting Agreement With Statement

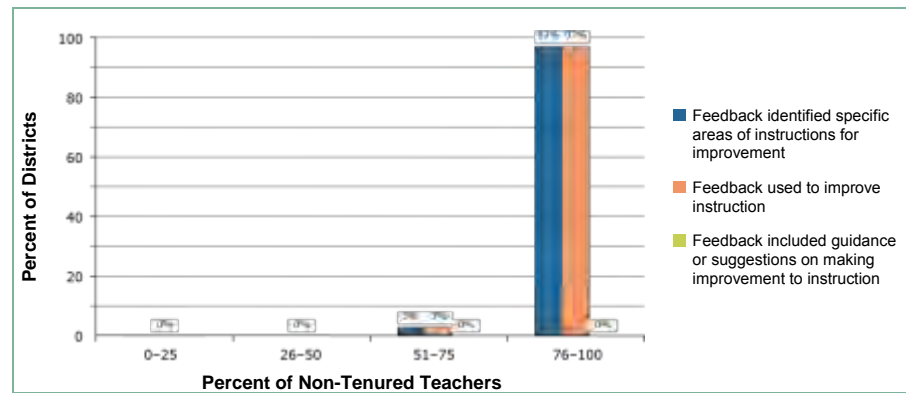


**Figure 7B.** Percent of Tenured Teachers Reporting Agreement With Statement

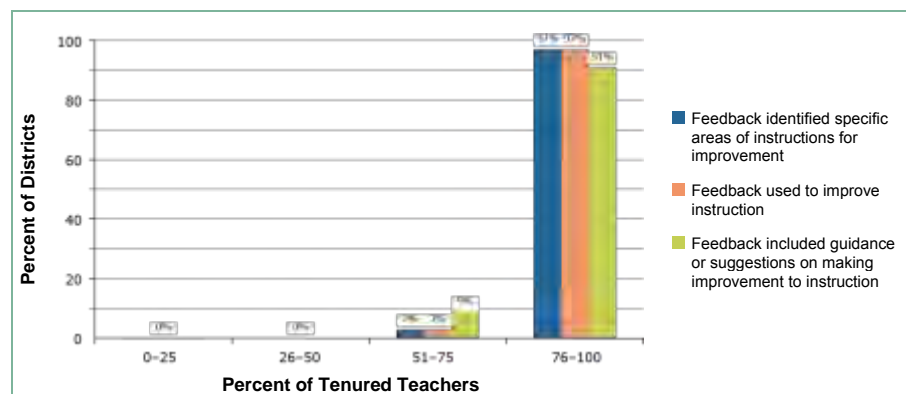




**Figure 7C.** Percent of Non-Tenured Teachers Reporting Agreement With Statement



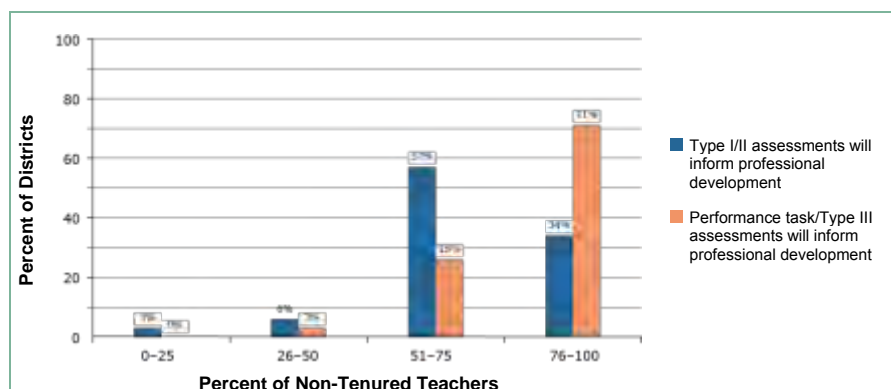
**Figure 7D.** Percent of Tenured Teachers Reporting Agreement With Statement



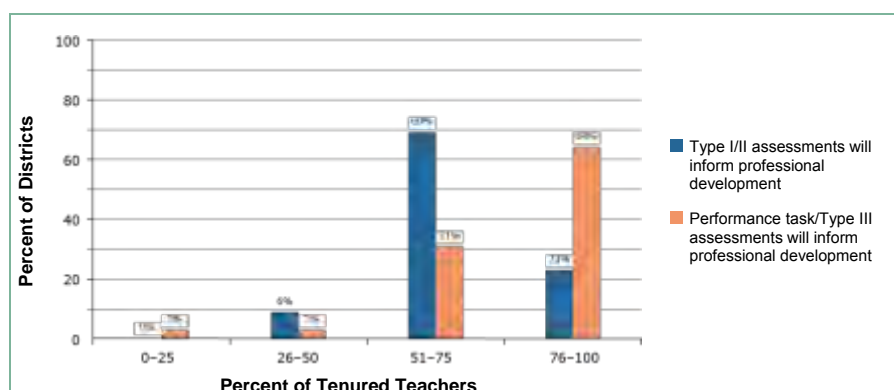
*Teacher Perceptions of the Impact of Student Growth Measures Based on Type I, II, and III Assessments on Professional Development*

- ▶ Compared to teacher observation perceptions, on average teachers were less inclined to agree that student growth measures based on Type I or II assessment results will inform professional development decisions. (See Figures 8a and 8b.)
  - In 91 percent of districts, 51 percent or more of non-tenured and tenured teachers agreed or strongly agreed that the information they get from the growth measure based on Type I and II assessments will inform professional development.
- ▶ On average, by district, teachers appear more likely to agree or strongly agree that student growth measures based on results from Type III assessments will inform professional development decisions.
  - In 97 percent of districts, 51 or more percent of non-tenured teachers reported that student growth based on Type III assessments will inform professional development. Similarly, in 94 percent of districts, 51 percent or more of tenured teachers reported that Type III assessments will inform professional development.

**Figure 8A.** Percent of Non-Tenured Teachers Reporting Agreement With Statement



**Figure 8B.** Percent of Tenured Teachers Reporting Agreement With Statement



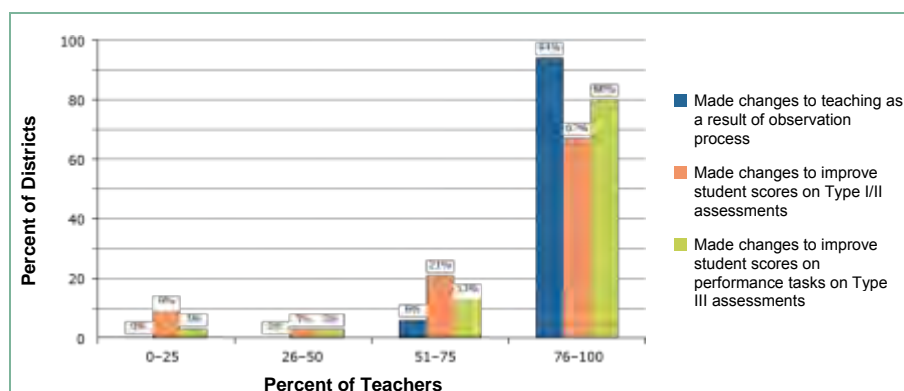
### Teacher Perceptions of the Impact of the Evaluation on Their Practice

Another key study question is how the implementation of PERA-compliant evaluation systems affected teacher practice. One key indicator of successful implementation of the teacher evaluation systems is whether teachers actually changed their practice as a result of the evaluation process. To address this question, the survey asked teachers about changes in their practice that were due to the evaluation process.

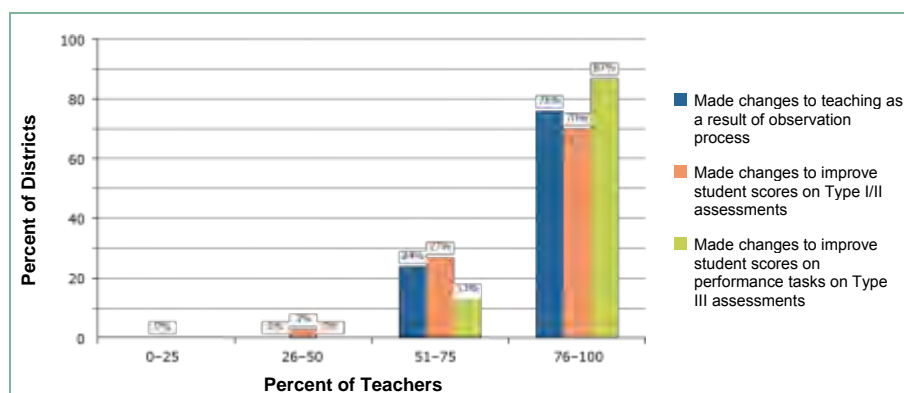
- ▶ In general, teachers agreed or strongly agreed that that they made changes to teaching as a result of the observation process, made changes to teaching to improve student assessment scores. (See Figures 9a and 9b.)
  - In 100 percent of districts, 51 percent or more of non-tenured and tenured teachers reported they agreed or strongly agreed that they made changes to teaching as a result of the observation process.
  - In 88 percent of districts, 51 percent or more of non-tenured teachers reported they agreed or strongly agreed that they made changes to improve student scores. In 97 percent of districts, 51 percent or more of tenured teachers reported they agreed or strongly agreed that they made changes to improve student scores on Type I/II assessments.
  - In 93 percent of districts, 51 percent or more of non-tenured teachers reported they agreed or strongly agreed that they made changes to improve student scores on a performance task. In every district, 51 percent or more of tenured teachers reported they agreed or strongly

agreed that they made changes to improve student scores on Type III assessments on performance tasks.

**Figure 9A.** Percent of Non-Tenured Teachers Within Districts Agreeing They Made Changes in Responses to the Evaluation



**Figure 9B.** Percent of Tenured Teachers Within Districts Agreeing They Made Changes in Responses to the Evaluation



- However, there were slight differences in the degree to which teachers agreed based on their tenure status.<sup>35</sup>

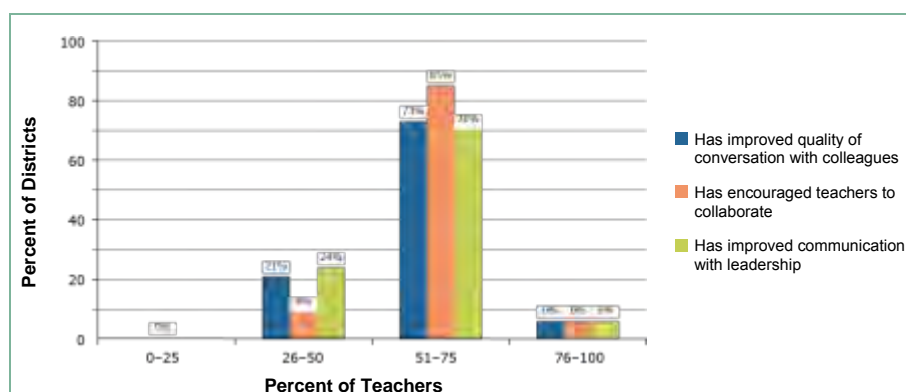
<sup>35</sup> Though both tenured and non-tenured teacher responses indicated support, the difference between tenured and non-tenured responses to the question of whether they made changes as a result of the observation process was statistically significant at the  $p > .001$  level (the difference was .276 and all responses were on a 4-point scale). Similarly, tenured teachers were slightly less likely to respond that they agreed with the statement that they

### Teacher Perceptions of Impact of Evaluation on School Culture

A major contributor to a positive school culture is the communication and collaboration among teachers and the communication between teachers and leadership. To assess the possible effect of the evaluation systems on communication and collaboration, the survey asked teachers several questions about these topics. The survey also asked teachers about the impact of these more rigorous evaluation processes on their stress level.

- ▶ In most districts, 51 percent or more of the teachers agreed or strongly agreed that that the evaluation process had positive impacts on collaboration and communication. (See Figure 10 below.)
  - In 79 percent of districts, 51 percent or more of teachers agreed or strongly agreed that the evaluation process improved the quality of conversations with colleagues.
  - In 91 percent of districts, 51 percent or more of teachers agreed or strongly agreed that the evaluation process encouraged teachers to collaborate.
  - In 75 percent of districts, 51 percent or more of teachers reported **agreement or strong agreement with the statement “the evaluation process has improved communication with leadership.”**

**Figure 10.** Percent of Teachers Within Districts Agreeing That Evaluation Has Improved Collaboration and Communication



changed their teaching to improve student scores. Notably, there was not a statistically significant difference between tenured and non-tenured teachers' responses to the question on whether they made changes to improve student scores on performance tasks.

- ▶ The evaluation systems appear to have created additional stress for teachers, particularly among the tenured teachers.
  - In 94 percent of districts, at least 51 percent of tenured teachers agreed or strongly agreed that the evaluation process created more stress.
  - In 85 percent of districts, at least 51 percent of non-tenured teachers agreed or strongly agreed that the evaluation process created more stress.
  - There was a difference in the level of agreement between tenured and non-tenured teachers.<sup>36</sup>

### Teacher Perceptions of Long-Term Impact of the Teacher Evaluation Process

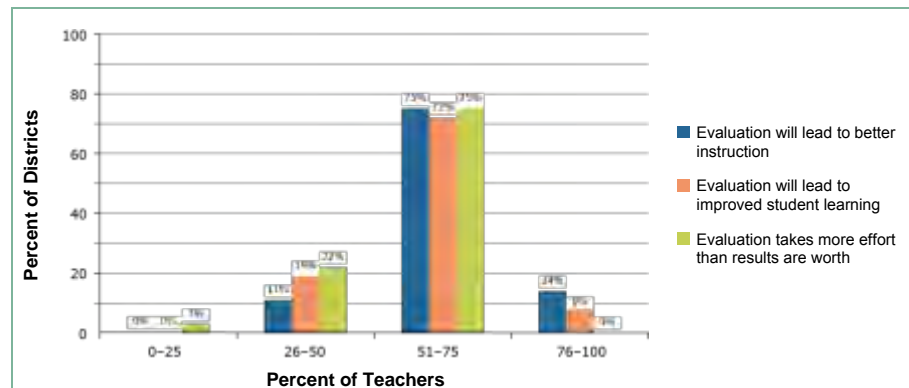
A central question is whether the implementation of the PERA-compliant teacher evaluation systems will have a long-term impact on teacher practice and on student learning. To begin to get a sense of potential long-term impacts of the teacher evaluation systems, the survey asked teachers questions about the long-term impact of teacher evaluation systems on teacher practice and student learning. Because the new systems appear to demand more work from teachers and evaluators, the survey also asked teachers whether they perceived the evaluation systems took more effort than the results were worth.

- ▶ In most districts, teachers were optimistic about the future impacts of the evaluation systems.
  - In the majority of districts, 51 percent or more of teachers reported agreeing or strongly agreeing that evaluation will result in better instruction and improved student performance. (See Figure 11 below.)
  - However, in relatively few districts did more than 75 percent of the responding teachers agree.
  - While teachers in many districts seemed to agree that the evaluation process will have positive impacts, in many districts, they appeared to doubt that the results are worth the extra effort. This is more likely to be the case for tenured teachers. In 40 percent of districts, 51 percent or more of non-tenured teachers agreed or strongly agreed that the evaluation system was more work than it was worth. In 88 percent of districts, 51 percent or more of tenured teachers agreed or strongly agreed with the same statement.

---

<sup>36</sup> The difference was statistically significant at the  $p < .001$  level.

**Figure 11.** Percent of Teachers Within Districts That Agreed That the Evaluation Process Will Improve Instruction and Student Learning



## Principal and Assistant Principal Perceptions and Understanding of the Teacher Evaluation Process

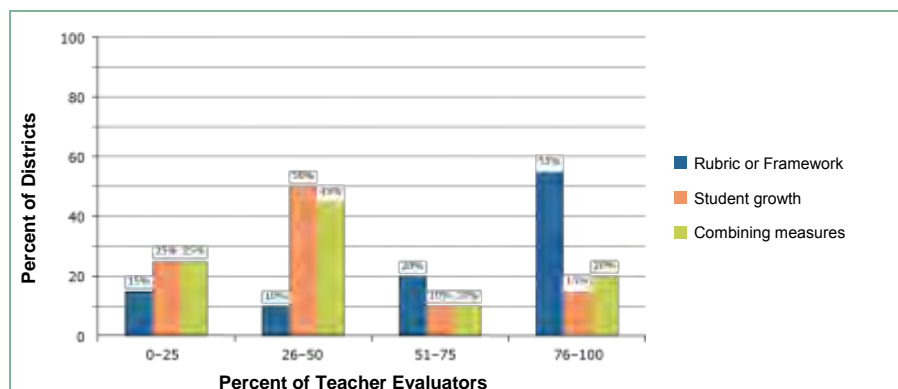
Principals and assistant principals are the typical evaluators of teachers. Therefore, in addition to surveying principals and assistant principals about their perceptions of the evaluation system, the study team also surveyed these school administrators about their understanding of the evaluation process, the time spent on evaluation activities, and the **evaluation process' impact** on teacher practice and school culture. The next section discusses these survey results and refers to the principals/assistant principals who responded as evaluators.

### Understanding of the Evaluation Process

To accurately conduct teacher evaluations, it is critical for evaluators to understand the components of the teacher evaluation systems. Understanding the evaluation process is necessary to establishing consistency in evaluator ratings across teachers. Evaluator understanding of the evaluation system also serves as another good indicator of the quality of a district's implementation of the teacher evaluation system. To gauge evaluator understanding of the evaluation systems, the survey examined evaluator responses to questions asking about their understanding of the evaluation rubric or framework, the student growth measure, and the method for combining the two measures into a final rating.

- ▶ In most districts, the majority of principals and assistant principals reported having strong understanding of the evaluation system practice rubrics. (See Figure 12 below.)
  - In 75 percent of the districts, 51 percent or more of the respondents rated their understanding as strong or very strong.
- ▶ These administrators were less likely to report a strong understanding of the growth scores and how practice and growth ratings were combined.
  - In 25 percent of the districts, 51 percent or more of the respondents rated their understanding of how the different assessment results combined to generate a growth score as strong or very strong. In addition, in 30 percent of the districts, 51 percent rated their understanding of how growth and practice scores combined as strong or very strong.

**Figure 12.** Percent of Principals/Assistant Principals Within Districts Reporting a Strong Understanding of Teacher Evaluation System Components





In most districts, the majority of evaluators rated their proficiency in applying the practice rubrics as strong or very strong.

- ▶ In 75 percent of the districts, 51 percent or more of principals and assistant principals rated their proficiency in understanding the differences between levels of the rubrics as strong or very strong.
- ▶ In 90 percent of the districts, 51 percent or more of principals and assistant principals rated their proficiency in determining what constitutes evidence related to the rubric components as strong or very strong.
- ▶ In 95 percent of the districts, 51 percent or more of principals and assistant principals rated their proficiency in recording evidence while observing as strong or very strong.
- ▶ In 85 percent of the districts, 51 percent or more of principals and assistant principals rated their proficiency in aligning evidence to the appropriate rubric component as strong or very strong.
- ▶ In 80 percent of the districts, 51 percent or more of principals and assistant principals rated their proficiency in determining observation ratings based on evidence as strong or very strong.
- ▶ In 75 percent of the districts, 51 percent or more of principals and assistant principals rated their proficiency in using the rubric to structure conversations with teachers as strong or very strong.

### *Teacher Evaluators Perception of Rater Agreement*

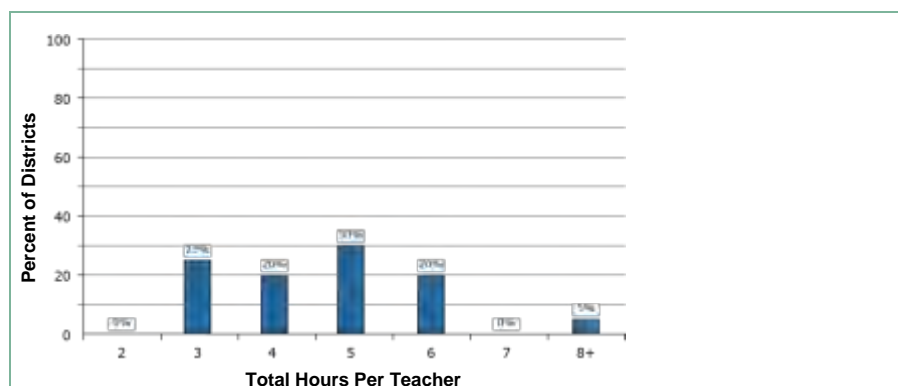
- ▶ In most districts, evaluators were moderately or very confident that if another evaluator rated the same teachers, those teachers would receive the same ratings.
  - In 85 percent of the districts, 51 percent or more evaluators responded that they were moderately or very confident.
  - However, in only 15 percent of the districts did 51 percent or more respond that they were very confident.

### Time Spent on Evaluation Activities

The survey asked evaluators to indicate how much time they spent conducting pre-conferences, observing teachers and deciding on ratings, and conducting post-conferences with each teacher they evaluated. Since PERA requires evaluators to conduct pre- and post-conferences and spend at least 40 minutes formally observing teachers, one would expect evaluators to spend a significant amount of time on evaluation activities. To address this question, the survey asked evaluators how much time they spent on teacher evaluations.

- ▶ Across districts, the time spent on the activities varied from three to more than eight hours. The median across the 20 districts was five hours. (See Figure 13 below.)
  - In the median district, evaluators spent one hour on pre-conference and one hour on post-conference activities.
  - In the median district, these evaluators spent two hours per teacher on observation and rating.

**Figure 13.** Average Total Number of Hours Spent on Teacher Evaluation Within Districts<sup>37</sup>



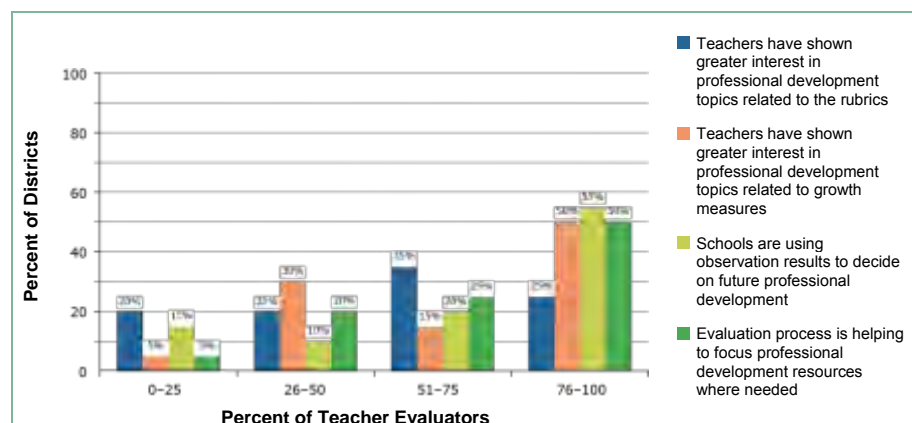
<sup>37</sup> Figure 13 reads: In 0% of districts with responses from principals and assistant principals evaluating teachers, the median total number of hours spent in evaluating each teacher was 2 hours. In 25% of the districts, the median was 3 hours. In 20% of the districts, the median was 4 hours. In 30% of the districts, the median was 5 hours. In 20% of the districts, the median was 6 hours. In 0% of the districts, the median was 7 hours, and in 5% of the districts, the median was 8 or more hours.

### Impact on Teacher Professional Development

Professional development is a primary mechanism for improving teacher practice, and one of the key questions of the study is whether the evaluation systems influence teacher professional development. As with teachers, we asked the evaluators whether the evaluation systems are influencing teacher professional development.

- ▶ In most districts, the majority of evaluators perceived that the evaluation system was affecting or will affect teacher professional development. (See Figure 14 below.)
  - In 60 percent of the districts, 51 percent or more of evaluators agreed or strongly agreed that teachers have shown a greater interest in professional development topics related to the professional practice standards and rubrics.
  - In 55 percent of the districts, 51 percent or more of evaluators agreed or strongly agreed that teachers have shown a greater interest in professional development topics related to improving achievement on their student growth measures.
  - In 75 percent of the districts, 51 percent or more of evaluators agreed or strongly agreed that they are using observation results to decide on the professional development topics the school will focus on in the future and that the evaluation process is helping the school focus professional development resources.

**Figure 14.** Percent of Teacher Evaluators Within Districts Agreeing That the Evaluation Process Has Influenced Teacher Professional Development

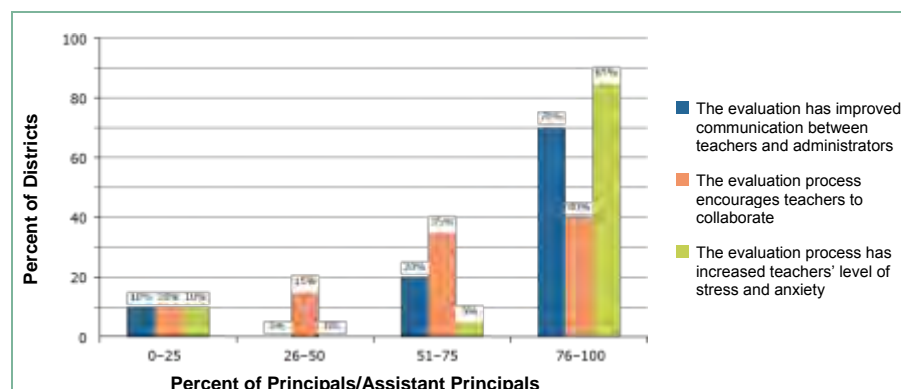


### Impact of School Culture

Similarly as for teachers, the survey asked evaluators about aspects of school culture, such as whether teacher evaluation systems were improving (1) teacher communication and collaboration and (2) teacher communication with leadership. The survey also asked whether the system created additional stress.

- ▶ While in almost all districts the majority of principals and assistant principals agreed that the evaluation improved communication between teachers and administrators and encouraged teachers to collaborate, in most districts the majority of principals and assistant principals also agreed that the evaluation process stressed teachers. (See Figure 15 below.)
  - In 90 percent of the districts, 51 percent or more of evaluators agreed or strongly agreed that the evaluation had improved communication between teachers and administrators at the school
  - In 75 percent of the districts, 51 percent or more of evaluators agreed or strongly agreed that the evaluation had encouraged teachers to collaborate.
  - In 90 percent of the districts, 51 percent or more of evaluators agreed or strongly agreed that the evaluation process had increased teacher stress and anxiety.

**Figure 15.** Percent of Principals/Assistant Principals Within Districts Agreeing That the Evaluation Process Has Affected Communication and Collaboration

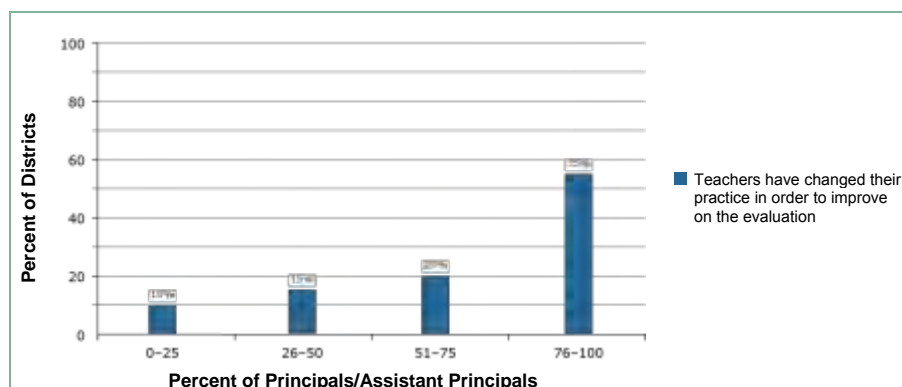


### Impact on Teaching Practice

As previously mentioned, a key question of the study is whether the evaluation systems affected teacher practice. Similarly as for teachers, we asked the evaluators about changes in teacher practice as a result of the evaluation process, as well as their perceptions of the future impact of teacher evaluation systems.

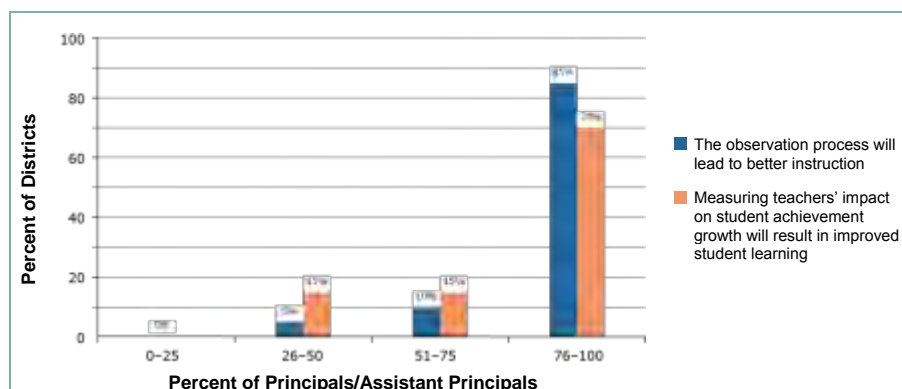
- ▶ Principals and assistant principals typically reported that about half the teachers they evaluated incorporated evaluator feedback into their teaching and made noticeable improvements in practice.
  - In the average district, 18 percent of the principals and assistant principals responded that all or most teachers incorporated evaluator feedback into their teaching, while 72 percent responded that about half of the teachers did, and 10 percent checked that few or none did.
  - In the average district, 31 percent of the principals and assistant principals responded that all or most teachers made noticeable improvements in their teaching; 63 percent responded that about half of the teachers did; and 6 percent checked that few or none did.
- ▶ In most districts, 51 percent or more of evaluators agreed or strongly agreed that teachers have changed the way they teach in order to perform better on the evaluation. (See Figure 16 below.)
  - While in 75 percent of these districts 51 percent agreed or strongly agreed that teachers changed their practice in order to perform better, three districts stood in contrast to the majority of districts. In two of these districts, none agreed with the same statement, and in the other district, less than 30 percent agreed.

**Figure 16.** Percent of Principals/Assistant Principals Within Districts Agreeing That the Evaluation Process Has Influenced Teacher Practice



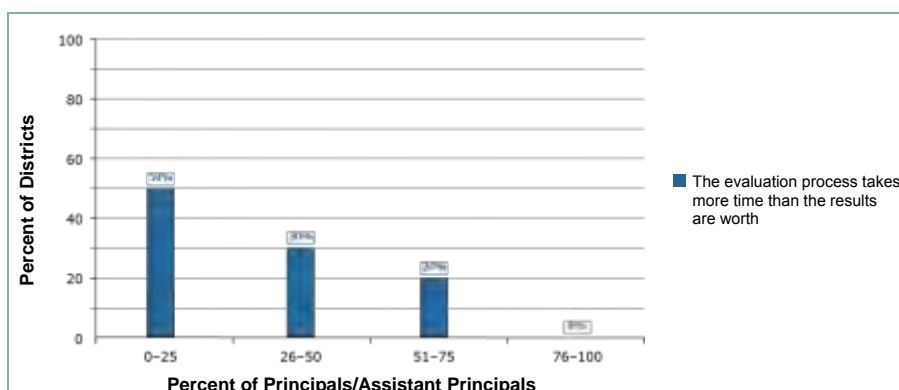
- ▶ In most districts, principals and assistant principals were optimistic about the long-term effects of the evaluation process. (See Figure 17 below.)
  - In 95 percent of the districts, 51 percent or more of evaluators agreed or strongly agreed that the observation process will lead to better instruction.
  - In 85 percent of the districts, 51 percent or more of evaluators agreed or strongly agreed that measuring teacher impact on student achievement growth will result in improved student learning.

**Figure 17.** Percent of Principals/Assistant Principals Within Districts Agreeing That Evaluation Process Will Have a Positive Effect on Instruction and Student Learning



- ▶ In most districts, a relatively low proportion of evaluators agreed or strongly agreed that the evaluation process took more time and effort than the results were worth, but in a few districts, the majority agreed. (See Figure 18 below.)
  - In 45 percent of the districts, 20 percent or fewer of evaluators agreed or strongly agreed that the evaluation process required more time and effort than it was worth.
  - However, in 20 percent of the districts, 51 percent or more agreed or strongly agreed that the process took more time and effort than the results were worth.

**Figure 18.** Percent of Principals/Assistant Principals Within Districts Who Agreed That the Teacher Evaluation Process Takes More Effort Than Results Are Worth



### Evaluator Responses to Open-Ended Questions Pertaining to Teacher Evaluation

The survey also provided space for respondents to add open-ended comments about the teacher evaluation process. Fifty-three evaluators in over 17 districts added comments.

- ▶ The most common theme in the comments (mentioned by 19 evaluators) was that the teacher evaluation process was very time intensive. However, some of these evaluators expressed additional opinions. Eleven noted that as well as being time intensive, the evaluation process added stress to various aspects of teaching. An

additional two evaluators believed that they lacked sufficient time to evaluate all staff members. On the other hand, six evaluators noted that while the process was time intensive, it contributed to teacher professional growth.

- ▶ A second common theme (mentioned by 10 evaluators) was that the process improved communication, collaborative discussion, and reflective thinking among staff.
- ▶ A third theme (mentioned by four evaluators) was that it was difficult to judge the impact of the new system at the time of the survey because it was new. Yet, the evaluators agreed it was likely to improve over time. Two additional respondents mentioned that the process improves each year.
- ▶ A fourth theme (mentioned by four evaluators) was the desire for additional training to increase **teachers'** understanding of the evaluation system.

### **Principal and Assistant Principal Perceptions of Their Own Evaluation Process**

We surveyed principals and assistant principals about their understanding of the principal evaluation process, the time spent on evaluation activities, and how the evaluation process influenced their practice.

#### Understanding of the Evaluation Process

In order for the evaluation process to have a positive impact on school administrator performance and be perceived as fair, those being evaluated have to understand the process and how performance ratings are made. To assess whether these conditions were present, the survey asked principals and assistant principals about their understanding of the components of the evaluation process.

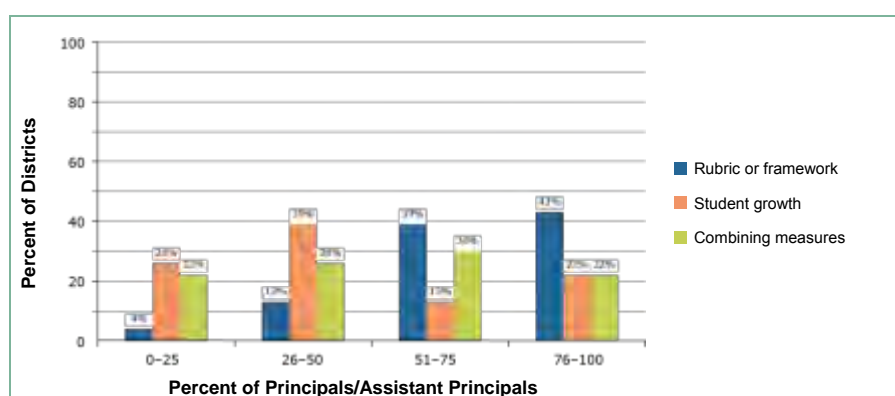
- ▶ In most districts, the majority of principals and assistant principals reported having a strong understanding of the evaluation system practice rubrics. (See Figure 19 below.)
  - In 83 percent of the districts, 51 percent or more of the principals or assistant principals rated their understanding as strong or very strong.
- ▶ Principals and assistant principals were less likely to report a strong understanding of the growth scores and how practice and growth ratings were combined. (See Figure 19 below.)
  - In 35 percent of the districts, 51 percent or more of principals and assistant principals rated their understanding of how the different



assessment results were combined to generate a growth score as strong or very strong.

- In 52 percent of the districts, 51 percent or more of principals and assistant principals rated their understanding of how growth and practice scores were combined as strong or very strong.

**Figure 19.** Percent of Principals/Assistant Principals Within Districts Reporting a Strong Understanding of Evaluation System Components



### Notification and Goal Setting

PERA requires that goals be set between principals/assistant principals and the evaluators and that principals/assistant principals receive a copy of the rubrics, have two observations, and receive formal written feedback within 10 days of the observation or site visit. Principals and assistant principals are also to complete a self-assessment. The survey asked if principals and assistant principals had experienced these activities.<sup>38</sup>

- Except for receiving feedback within 10 days of an observation, in most districts all principals and assistant principals responded that these activities had taken place.
  - In 96 percent of the districts, all principals and assistant principals responded that they had received rubrics.

<sup>38</sup> Note that for CPS, the survey asked if principals had yet experienced one observation or site visit, rather than two, because the CPS timeline required evaluations to be completed by July 1, 2014, rather than in March 2014, as for other districts.

- In 87 percent of the districts, all principals and assistant principals reported that they had met with their evaluator to set goals.
- In 78 percent of the districts, all principals and assistant principals noted that they had completed a self-assessment.
- In 74 percent of the districts, all principals and assistant principals checked that their evaluator had made one (in CPS) or two (in other districts) site visits (by the time of survey administration).
- However, in only 48 percent of the districts did all principals and assistant principals report receiving feedback within 10 days of the visit.

### Time Spent on Evaluation Activities

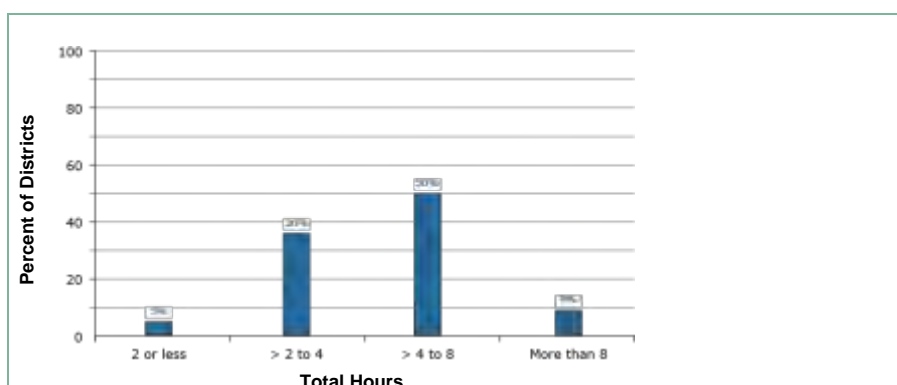
Another factor that affects the acceptance and long-term viability of evaluation systems is the time required to carry out the activities. We asked principals and assistant principals<sup>39</sup> about the number of hours (up to the survey administration) spent on goal setting, completing the self-evaluation, preparing for observations, and reviewing and discussing feedback with their evaluators.

- ▶ The most common median number of hours spent by principals and assistant principals on all these activities was between four to eight hours. (See Figure 20 below.)
  - In 50 percent of the districts, the median number of hours spent was more than four but less than eight.

---

<sup>39</sup> The survey did not ask this question of CPS principals due to survey space limitations.

**Figure 20.** Average Total Number of Hours Spent by Principals and Assistant Principals on Their Own Evaluation<sup>40</sup>



#### Principal and Assistant Principal Perceptions Related to Credibility of Their Evaluator and Fairness of the Rubric

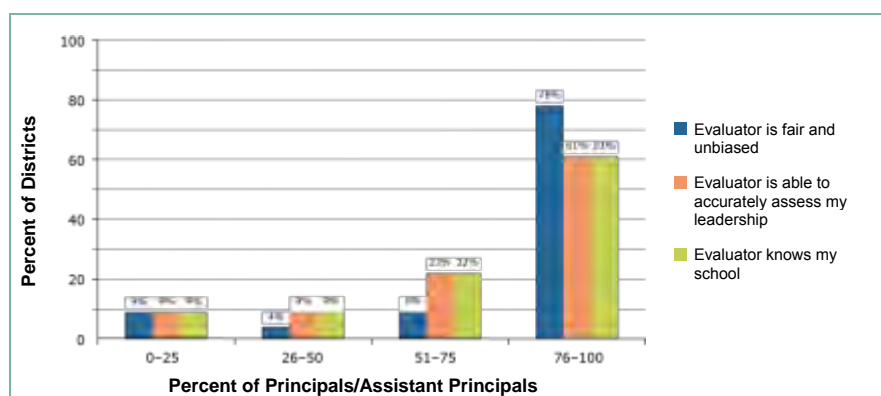
The acceptance and sustainability of evaluation systems is likely also affected by whether those receiving evaluations perceive the process as fair. The survey asked principals and assistant principals about several aspects of fairness, including the credibility of the evaluator, **the consistency of evaluators' ratings** across principals, and the overall fairness of the leadership practice and student growth components.

- ▶ In most districts, the majority of principals and assistant principals responding perceived the evaluator as credible and fair. (See Figure 21 below.)
  - In 78 percent of the districts, 51 percent or more of the principals and assistant principals responded “some” or “to a great extent” when asked how true it was that the evaluator was fair and unbiased.
  - In 87 percent of the districts, 51 percent or more of the principals and assistant principals reported “some” or “to a great extent” when asked how true it was that the evaluator was able to accurately assess professional practice.

<sup>40</sup> Figure 20 reads: In 5% of the districts with responses from principals and assistant principals about their own evaluation, the average time they reported spending on their own evaluation was 2 or fewer hours. In 36% of the districts, the average time reported was greater than 2 up through 4 hours. In 50% of the districts, the average time reported was greater than 4 through 8 hours. And in 9% of the districts, the average time reported was greater than 8 hours.

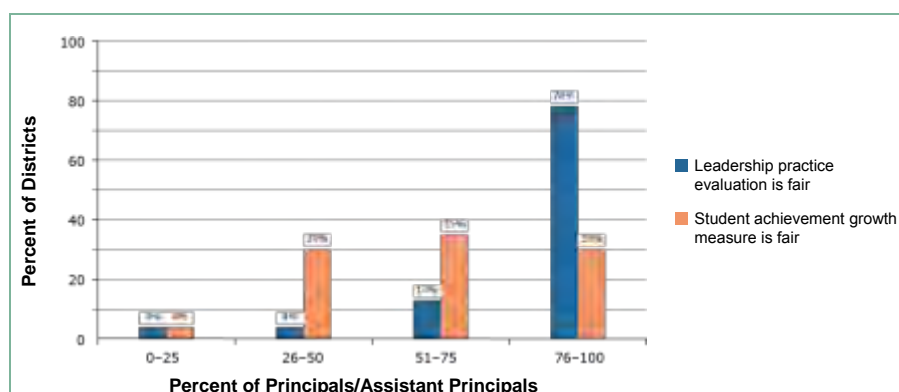
- In 83 percent of the districts, 51 percent or more of the principals and assistant principals checked “some” or “to a great extent” when asked how true it was that the evaluator knows what was going on in the principal’s/assistant principal’s school.

**Figure 21.** Percent of Principals/Assistant Principals Within Districts Reporting Evaluators as Credible and Fair



- ▶ In most districts, the majority of principals and assistant principals did not perceive that evaluators rated them inconsistently.
  - In 26 percent of the districts, 51 percent or more of the principals and assistant principals agreed or strongly agreed that the evaluation ratings school leaders receive depend more on their evaluator than on their professional practice.
- ▶ In most districts, the majority of principals and assistant principals agreed that the leadership practice part of the evaluation was fair. (See Figure 22 below.)
  - In 91 percent of the districts, 51 percent or more of the principals and assistant principals agreed or strongly agreed that the leadership practice part of the evaluation process was fair.
- ▶ In fewer districts did the majority of principals and assistant principals agree that the student growth measurers were fair. (See Figure 22 below.)
  - In 65 percent of the districts, 51 percent or more of the principals and assistant principals agreed or strongly agreed that the student achievement growth measures used to evaluate their performance were fair.

**Figure 22.** Percent of Principals/Assistant Principals Within Districts Agreeing That Components of Their Evaluation Process Were Fair



### Impact of PERA Evaluation on Professional Development

Noted above, one of the key questions of the study was to better understand the impact of the implementation of PERA-compliant evaluation systems on professional development. To address this question, the survey asked principals and assistant principals about the feedback they received and how the evaluation influenced their professional development choices.

- ▶ In 96 percent of the districts, 51 percent or more of the principals and assistant principals reported receiving feedback from their evaluator.
- ▶ In 73 percent of the districts,<sup>41</sup> 51 percent or more of principals and assistant principals who received feedback rated it as useful or very useful.

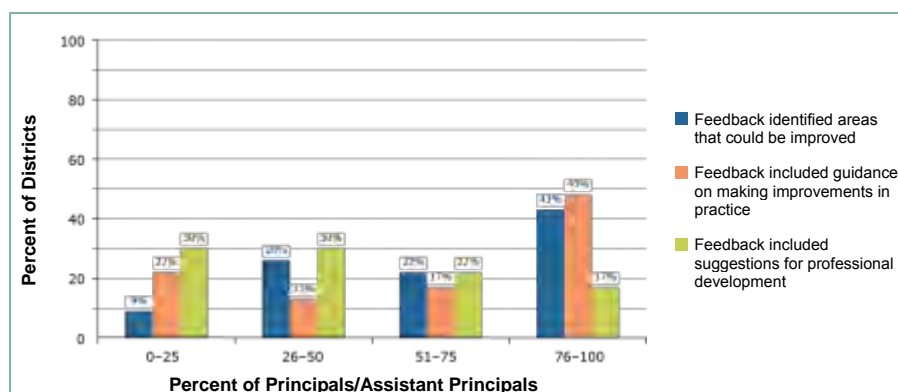
Across all 23 districts, the survey asked principals and assistant principals about specific feedback received on leadership practice.

- ▶ In most districts (65 percent), 51 percent or more of principals and assistant principals responded that it was true or mostly true that the feedback they received identified specific areas of practice that they could improve, including guidance on how to make improvements.
- ▶ However, in only 39 percent of the districts, 51 percent or more of principals and assistant principals responded that it was true or mostly true that the feedback they received included suggestions for specific professional development that could help improve performance.

<sup>41</sup> The survey did not ask this question of CPS principals due to space limitations on the CPS survey.

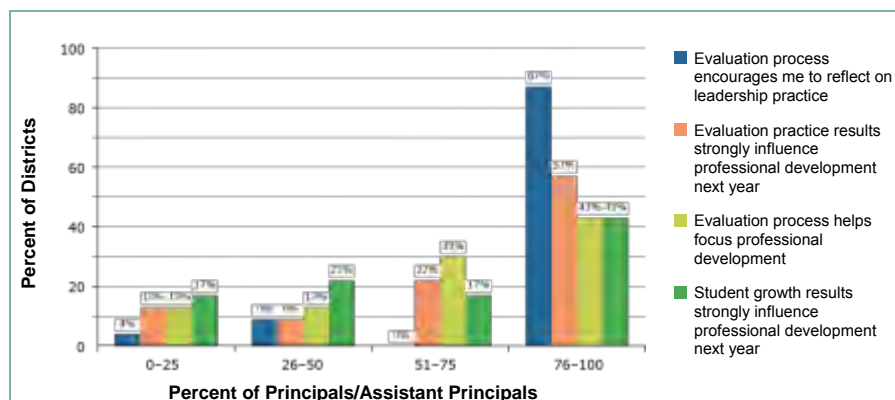
Figure 23 below shows more information about the percentage of districts in which various percentages of principals responded that these statements about feedback were true or mostly true.

**Figure 23.** Percent of Principals/Assistant Principals Within Districts Reporting Statements About Feedback Were True or Mostly True



- ▶ In most districts, a majority of principals and assistant principals agreed or strongly agreed that the evaluation process affected their professional development efforts. (See Figure 24 below.)
  - In 87 percent of the districts, 51 percent or more of principals and assistant principals agreed or strongly agreed that the evaluation process encouraged them to reflect on their leadership practice.
  - In 78 percent of the districts, 51 percent or more of principals and assistant principals agreed or strongly agreed that their evaluation results would strongly influence their professional development activities next year.
  - In 74 percent of the districts, 51 percent or more of principals and assistant principals agreed or strongly agreed that the evaluation process is helping them focus their professional development efforts where most needed.
  - In 61 percent of the districts, 51 percent or more of principals and assistant principals agreed or strongly agreed that their student growth results will strongly influence their professional development activities next year.

**Figure 24.** Percent of Principals/Assistant Principals Within Districts Agreeing That the Evaluation Process Affected Their Professional Development

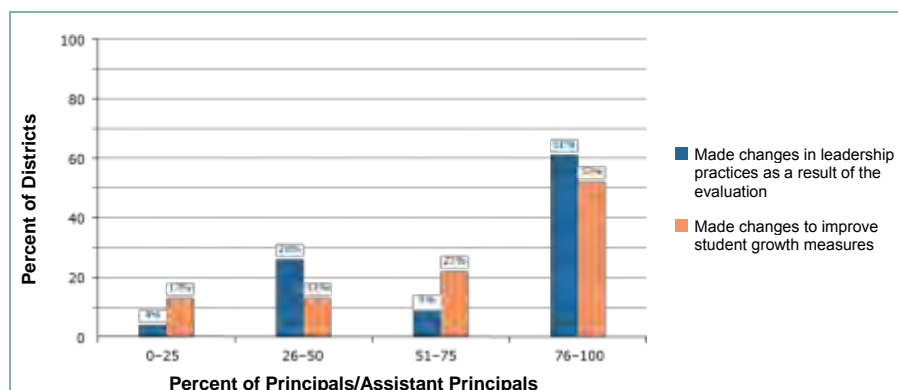


### Impact of PERA Evaluation on School Leader Practice

Another key question of the study was how the implementation of PERA-compliant evaluation systems affected principal behavior or practice. To address this question, the survey asked principals and assistant principals about the changes they made to their practice as a result of the evaluation.

- ▶ In most districts, the majority of principals and assistant principals reported changing behavior as a result of the evaluation process. (See Figure 25 below.)
  - In 70 percent of the districts, 51 percent or more of principals and assistant principals agreed or strongly agreed that they made changes in their leadership practice as a result of their evaluation.
  - In 74 percent of the districts, 51 percent or more of principals and assistant principals agreed or strongly agreed that they changed their behavior to improve their student growth measures.

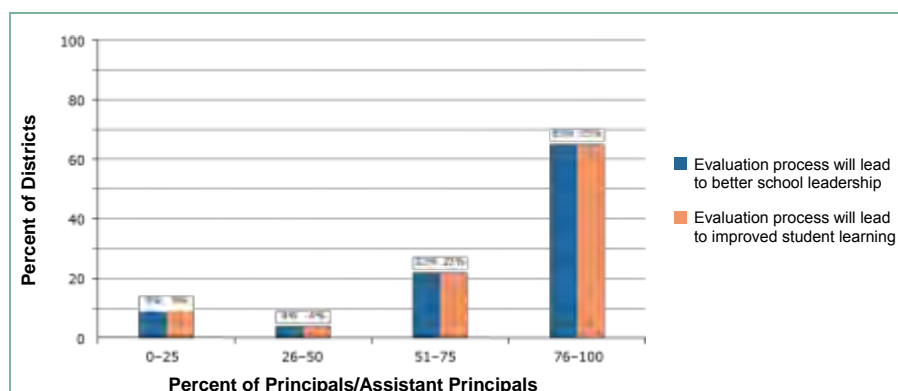
**Figure 25.** Percent of Principals/Assistant Principals Within Districts Agreeing That the Evaluation Process Affected Their Practice



In addition, the survey asked principals and assistant principals about their perceptions of the future impact of the principal evaluation process.

- ▶ In most districts, most principals and assistant principals agreed that the evaluation process will lead to better school leadership and improved student achievement. (See Figure 26 below.)
  - In 87 percent of the districts, 51 percent or more of principals and assistant principals agreed or strongly agreed that the evaluation process will lead to better school leadership and improved student achievement.

**Figure 26.** Percent of Principals/Assistant Principals Within Districts Agreeing That the Evaluation Process Will Have a Positive Effect on Leadership Practice and Student Learning





The survey asked principals and assistant principals about whether the benefits of the more rigorous and comprehensive PERA evaluation system were worth the time and effort involved.

- ▶ In most districts, the majority of principals and assistant principals expressed concerns about the level of time and effort involved.
  - In 70 percent of the districts, 51 percent or more of principals and assistant principals agreed or strongly agreed that the evaluation process takes more time and effort than the results are worth.

#### Principal and Assistant Principal Open-Ended Responses About Administrator Evaluation System

The survey provided space for respondents to add open-ended comments about the principal evaluation process. Forty-three principals and assistant principals from 15 districts added comments. The bullets below summarize the most common themes expressed in these comments.

- ▶ Eight principals/assistant principals commented that the process was worthwhile and produced positive results, with an additional two commenting that they had experienced positive results despite the intensive time commitment.
- ▶ Seven principals/assistant principals believed that the growth model should better reflect the various duties principals perform rather than basing growth on actions that principals cannot control. Likewise, two principals/assistant principals believed that the process will always be subjective.
- ▶ Many principals and assistant principals noted various frustrations, including five who believed that the evaluation was too time consuming and strenuous.
- ▶ Four principals/assistant principals believed the process was biased and too formulaic.
- ▶ Three principals/assistant principals agreed that collaboration between the principal/assistant principal and the evaluator must increase.
- ▶ Additional comments regarding evaluators included three principals/assistant principals who believed that their evaluator must improve his/her own performance/role.

## Principal Evaluator Perceptions of the Principal Evaluation Process

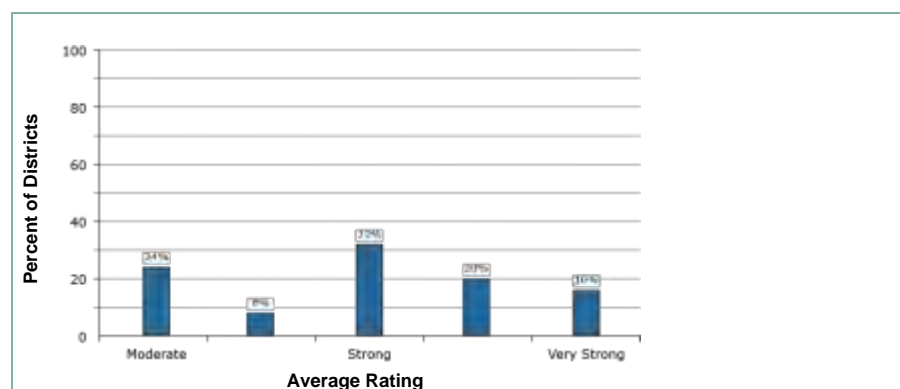
In addition to surveying principals and assistant principals about their perceptions of the evaluation system, the research team also surveyed their evaluators: superintendents, assistant superintendents, other central office administrators, and occasionally retired principals. The survey asked evaluators about their understanding of the principal evaluation process, the time spent on evaluation activities, and the impact of the system on principals' leadership practice.

### Understanding of the Evaluation Process

Like principals evaluated by the new PERA evaluation systems, principal evaluators must understand the evaluation process. To this end, the survey asked evaluators about their overall level of understanding of the evaluation system, as well as their understanding of the specific evaluation components (i.e., practice and growth).

- ▶ In most districts, evaluators rated their overall understanding of the evaluation system as strong or very strong. In only 32 percent of the districts did evaluators rate their understanding below the “strong” level (Figure 27).

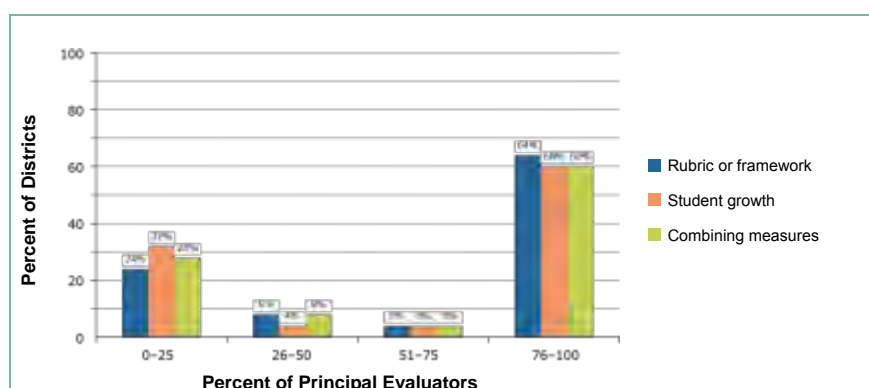
**Figure 27.** Percent of Principal Evaluators Within Districts Rating Understanding of Principal Evaluation System as Strong<sup>42</sup>



<sup>42</sup> Figure 27 reads: In 24% of the districts with responses from principals' evaluators, the average rating of these evaluators' understanding of the overall evaluation system was moderate. In 8% of the districts, the average rating was between moderate and strong. In 32% of the districts, the average rating was strong. In 20% of the districts, the average rating was between strong and very strong. And in 16% of the districts, the average rating was very strong.

- ▶ In most districts, 75 percent or more of the principal evaluators rated their understanding of the leadership practice rubric or framework, combining different assessments to create a student growth score, and combining the leadership practice and growth scores to determine a final summative rating as strong or very strong. (See Figure 28 below.)
  - However, in a substantial proportion of districts, 25 percent or fewer of principal evaluators rated their understanding as strong or very strong.

**Figure 28.** Percent of Principal Evaluators Within Districts Rating Understanding of Components of Principal Evaluation System as Strong



- ▶ Districts varied in the extent to which principal evaluators received training on the principal evaluation system beyond the required Growth Through Learning modules.
  - In approximately one-third of districts, all principal evaluators reported receiving additional training. However, in the next one-third of districts, between 25 and 75 percent of evaluators reported receiving additional training, while in the final one-third of districts, none of the evaluators reported receiving additional training.
  - Across the districts that provided additional training, the average number of hours was 8.4.
  - There is a small positive relationship between reporting receipt of additional training and understanding of the evaluation system components, suggesting that extra training leads to better understanding of the system.

### Principal Evaluator Perceptions on Rater Agreement

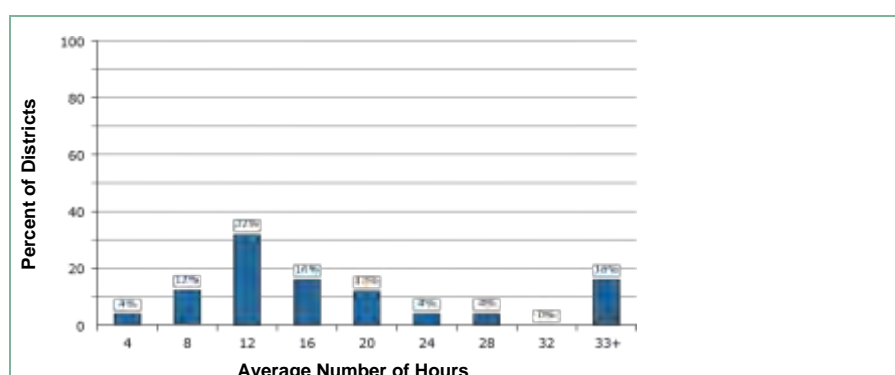
- ▶ Principal evaluators were confident that other evaluators rating the principals they observed would have provided similar ratings.
  - In 19 of the 21 districts with more than one principal evaluator, all evaluators responded that they were moderately or very confident that if another evaluator were to observe the principals they rated, those principals would receive the same rating. In the other two districts, 70 percent and 75 percent of evaluators responded that they were moderately or very confident, respectively.

### Time Spent on Evaluation Activities

PERA's more rigorous requirements for principal evaluation may require district evaluators to spend a substantial amount of time evaluating each principal. The survey asked principal evaluators to report on the average number of hours they spent preparing for the evaluation, collecting evidence, making ratings, and discussing them with each principal.

- ▶ The average hours spent varied substantially across districts, ranging from 4 to 33 hours or more. The most common average reported within districts was 12 hours per principal evaluated. (See Figure 29 below.)

**Figure 29.** Average Number of Hours Spent by Principal Evaluators Evaluating Each Principal<sup>43</sup>



### Impact of PERA Evaluation on Staff Development

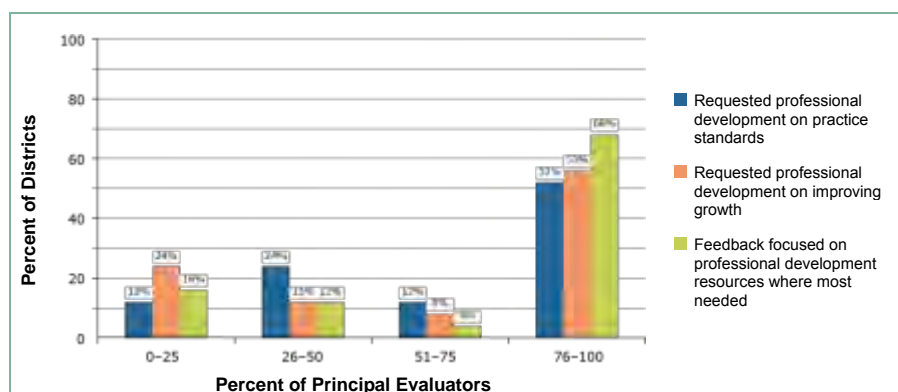
As previously mentioned, one of the key questions of the study is the impact of the PERA-compliant evaluation system implementation on staff development. The survey asked principal evaluators about the impact of the principal evaluation process on principal professional development.

- ▶ In most districts, principal evaluators agreed or strongly agreed that the professional practice rubric was useful for professional development.
  - In 84 percent of districts, all evaluators agreed or strongly agreed that the rubric was a useful tool for identifying which principals need more professional development.
  - In 88 percent of districts, all evaluators agreed or strongly agreed that the rubric was a useful tool for providing targeted support for principals.
  - In 84 percent of districts, all evaluators agreed or strongly agreed that the rubric improved the quality of conversations about school leadership with principals.

<sup>43</sup> Figure 29 reads: In 4% of the districts with responses from principals' evaluators, the average time these evaluators reported spending on each principal was 4 hours or less. In 12% of the districts, the average time reported was more than 4 through 8 hours. In 32% of the districts, the average time reported was more than 8 through 12 hours. In 16% of the districts, the average time reported was more than 12 through 16 hours. In 12% of the districts, the average time reported was more than 16 through 20 hours. In 4% of the districts, the average time reported was more than 20 through 24 hours. In 4% of the districts, the average time reported was more than 24 through 28 hours. In 0% of the districts, the average time reported was more than 28 through 32 hours, and in 16% of the districts, the average time was 33 hours or more.

- ▶ In most districts (though fewer compared to responses for practice rubrics), principal evaluators agreed or strongly agreed that the student growth measures were useful for professional development.
  - In 64 percent of the districts, all evaluators agreed or strongly agreed that the growth measures were useful for identifying which principals need more professional development, providing targeted support for principals, and improving the quality of conversations about school leadership.
- ▶ Principal evaluator responses confirmed the finding from the principals and assistants that the evaluation process influenced principal professional development efforts. (See Figure 30 below.)
  - In over 50 percent of the districts, 75 percent or more of the evaluators agreed or strongly agreed that principals requested professional development on topics related to the leadership standards and improving their student growth measures, showing that the evaluation process is helping focus professional development resources where they are needed most.

**Figure 30.** Percent of Principal Evaluators Within Districts Agreeing That the Evaluation Process Affected Principal’s Professional Development

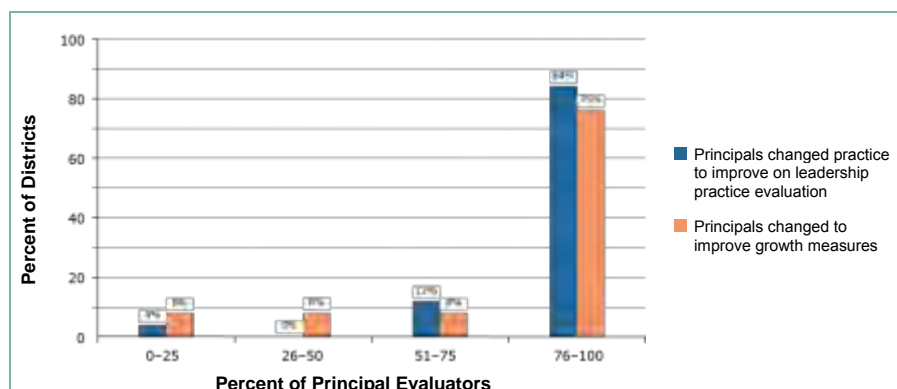


### Impact on Principal Practice

To assess the impact of the evaluation process on principal practice, the survey asked principal evaluators about potential impacts of the evaluation practice and growth components on principal behavior.

- ▶ In most districts, a large majority of evaluators agreed that the evaluation process affected principal behavior. (See Figure 31.)
  - In over 75 percent of the districts, 76 percent or more of the respondents agreed or strongly agreed that the principals they evaluated changed behavior to improve their leadership practice in order to increase student growth.

**Figure 31.** Percent of Principal Evaluators Within Districts Agreeing That the Evaluation Process Affected Principal Practice

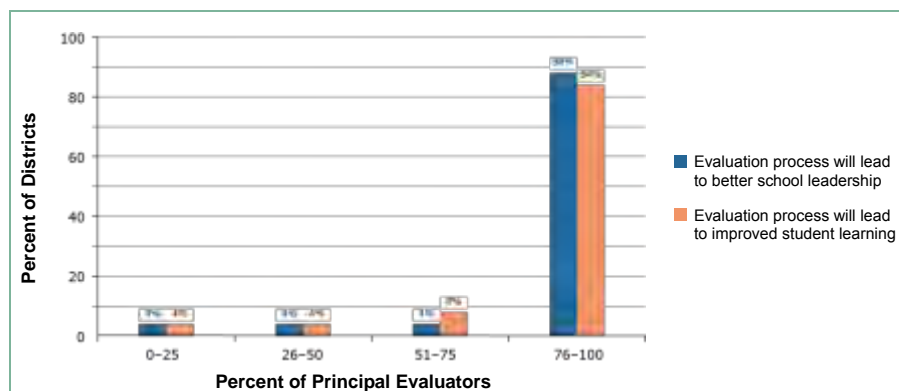


### Potential Long-Term Impact

Like principals, the survey asked principal evaluators about their perceptions about the potential future implications of the evaluation process.

- ▶ In most districts, the majority of principal evaluators agreed that the evaluation process will lead to better school leadership and improved student achievement. (See Figure 32.)
  - In over 80 percent of the districts, 76 percent or more of evaluators agreed or strongly agreed.

**Figure 32.** Percent of Principal Evaluators Within Districts Agreeing That the Evaluation Process Will Improve Leadership Practice and Student Learning

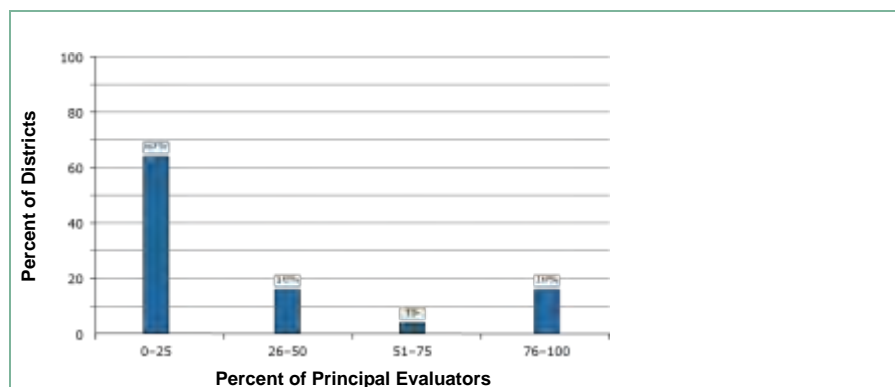


As asked of other survey participants, the survey asked principal evaluators about whether the benefits of the more rigorous and comprehensive PERA evaluation were worth the time and effort.

- ▶ Overall, principal evaluators were less likely than principals to report that the evaluation process takes more time and effort than the results are worth. (See Figure 33.)
  - In only 28 percent of the districts did 51 percent or more of evaluators agree or strongly agree that the evaluation process takes more time and effort than the results are worth.
  - However, in 64 percent of the districts, 51 percent or more of the evaluators agreed or strongly agreed that the evaluation process increased principal stress and anxiety.



**Figure 33.** Percent of Principals Evaluators Within Districts Agreeing That Principal Evaluation Process Takes More Time and Effort Than the Results Are Worth



### Principal Evaluator Responses to Open-Ended Questions on the Evaluation Process

As for other survey respondents, the survey provided space for principal evaluators to add open-ended comments about the principal evaluation process. Forty evaluators from 19 districts provided comments. The most common themes are summarized below

- ▶ The majority of evaluators commented that the process was very time consuming. Eight evaluators mentioned despite the time, the process yielded positive results, including increasing important discussion that helped educators to reflect on their practice and make improvements.
- ▶ In addition to comments regarding time commitment, six evaluators believed that the process was redundant and required a high level of organization and dialog in order to be efficient and successful.
- ▶ Three evaluators wrote that the new system increased district professional development efforts. However, three evaluators also noted the process did not encourage principals in specific areas of growth.

## **Additional Findings**

The following section discusses some additional patterns of responses across survey items and the different types of respondents (i.e., teachers, principals and assistant principals, and evaluators of principals).

### *Teachers*

- ▶ Teachers reported lower levels of understanding and fairness of the student growth component versus the professional practice component. While in the vast majority of districts, most teachers (76 percent and above) perceived the evaluation of professional practice as credible and fair, teacher responses to the fairness of the growth component were mixed.
- ▶ Teachers reported high regard for the feedback received after the classroom observation, whereas there was less of an indication that student growth would be used to inform professional development, though teachers indicated student growth data based on Type III assessments were more likely to inform professional development than student growth data based on Types I and II assessments.
- ▶ In a larger percentage of districts, non-tenured teachers reported that they made changes as a result of the observation process as opposed to the growth measures.

### *Teacher Evaluators*

- ▶ Similar to the teacher responses, teacher evaluators were less likely to report a strong understanding of the growth measures than the practice rubrics.

### *Principals and Assistant Principals*

- ▶ Principals and assistant principals were more inclined to rate their understanding of the practice rubric as strong versus the student growth calculation.
- ▶ Principals and assistant principals were more inclined to rate the leadership practice evaluation as fair versus the student growth measure.
- ▶ Principals and assistant principals were less likely to agree that the student growth measures would influence professional development in the following year versus the professional practice rubrics.
- ▶ Principals and assistants agreed that they would change their practice to improve their scores on the practice evaluation and increase student growth.

### *Principal Evaluators*

- ▶ Principal evaluators rated their understanding of the practice rubric and growth calculation at similar levels.
- ▶ Although principal evaluators agreed that both the practice rubric and growth were useful for professional development, fewer agreed student growth measures were useful.
- ▶ Principal evaluators agreed or strongly agreed that principals changed their behavior to improve both practice evaluation ratings and growth scores.

### Teachers Compared to Teacher Evaluators (Principals and Assistant Principals)

Teacher and teacher evaluator survey responses substantiated each other on understanding of the evaluation process, evaluator proficiency with rubrics, usefulness of the rubrics with improving instruction, and impact on professional development. However, there were some differences in responses across teachers and evaluators on rater consistency (specifically the importance of the evaluator on instructional practice ratings), communication between teachers themselves and leadership, and on the worthwhileness of the evaluation process. There also were some differences between teachers and evaluators on the impact of teaching practices and long-term effects of the evaluation systems.

- ▶ Both teacher and teacher evaluators reported high levels of agreement on understanding the teacher rubrics but low levels of agreement on understanding the growth measures.
- ▶ Evaluators rated their proficiency applying the practice rubrics as strong, which substantiates teacher responses on the credibility of their evaluators.
- ▶ However, while in most districts a majority of evaluators reported that they were confident that another evaluator would rate teachers at a similar rating, in a strong majority of districts, a majority of teachers reported that the observation ratings depended more on the evaluator than on teaching practices.
- ▶ Both teachers and evaluators indicated that the observation process affected teacher practice.
- ▶ In most districts, a majority of both teachers and evaluators perceived that the evaluation system affected teacher professional development or will guide future professional development.
- ▶ While in most districts the majority of evaluators reported that the evaluation improved communication between teachers and leadership and encouraged teachers

to collaborate, in fewer districts did as high a percentage of teachers agree that the evaluation system improved collaboration with their colleagues or communication with leadership.

- ▶ In a majority of districts, most teachers indicated they made changes to their teaching as a result of the observation process; yet, evaluators only perceived that about half of the teachers they evaluated incorporated evaluator feedback into teaching and made improvements.
- ▶ Evaluators appeared more inclined than teachers to report that the evaluation process was worthwhile. For example, in only 20 percent of districts, 51 percent or more of the evaluators agreed that the process required more time and effort than it was worth. Comparatively, both tenured and non-tenured teachers were more likely to report that the process required more time and effort than it was worth. In 89 percent of districts, 51 percent or more of tenured teachers agreed that the process required more time and effort than it was worth, and in 40 percent of districts, 51 percent or more of non-tenured teachers agreed or strongly agreed with the same statement.

### Principals Compared to Principal Evaluators

For most topics covered on the surveys, principal and principal evaluator responses validated each other. However, there were some differences in the reported levels of understanding and some slight differences in the perception of potential long-term effects and overall worth of the evaluation process.

- ▶ **There were some differences between principals and principal evaluators' responses** regarding understanding the evaluation process. Principals reported high levels of understanding for the professional practice rubric, relatively low levels of agreement for understanding student growth, and mixed responses for combining measures. However, principal evaluators reported high levels of understanding of the rubric, the student growth measures, and how the elements combined. Notably, however, there were a substantial proportion of districts in which less than 25 percent of principal evaluators rated their understanding of the evaluation components as strong.
- ▶ Both principals and evaluators indicated that the evaluation process influenced professional development.
- ▶ Both principals and evaluators noted that the evaluation process affected leadership practices, though evaluators were more optimistic about it.
- ▶ Both principals and evaluators reported that the evaluation process will have a long-term impact and will lead to better school leadership and improved student learning, though principal evaluators were more optimistic about it.

- ▶ Evaluators were more likely than principals to indicate that the evaluation process was worthwhile.

### Evaluators Compared to Evaluatees

There were some similarities between principal and teacher responses to their own evaluation system when compared to the responses of their evaluators. Notably, principals' and assistant principals' experience as teacher evaluators did not always appear to affect their perceptions of the credibility of their own evaluators.

- ▶ Teachers, teacher evaluators, principals, and principal evaluators were likely to agree with statements pertaining to the credibility of the evaluators. Evaluators tended to be confident that their practice ratings were consistent with those of other raters, but evaluatees tended to agree that ratings depended more on the evaluator than on their practice.
- ▶ Teacher and principal evaluators perceived the evaluation process as more worthwhile than teachers and principals.

### Teachers

- ▶ On the teacher survey, in a majority of districts, a majority of tenured and non-tenured teachers gave favorable responses across most of the survey items by topic (e.g., understanding, fairness, impact on professional development).
- ▶ While in a large percentage of districts most teachers reported that they received information regarding the evaluation systems, in fewer districts did the majority report a strong understanding the evaluation system components.
- ▶ A strong majority of tenured and non-tenured teachers reported that their evaluator was credible, yet in a majority of districts most teachers still reported that their observation rating depended more on the evaluator than on the instructional practices.
- ▶ A majority of teachers reported a strong consensus that the evaluation system would affect their practice, but nevertheless still reported mixed responses on the evaluation system leading to better instruction and to improved student learning. Notably, teachers also tended to agree that the evaluation systems required more time and effort than they were worth.

### *Teacher Evaluators*

- ▶ A majority of evaluators reported a strong understanding of the evaluation system, rated their proficiency in the practice rubric as strong, and were confident that there was consistency across evaluators.
- ▶ A majority of evaluators perceived that the rubrics were useful for helping teachers improve and affected teacher professional development.
- ▶ While evaluators reported that only about half of the teachers evaluated made changes to teaching practices, they were more optimistic than teachers about the long-run effects of the evaluation process.

### *Principals and Assistant Principals Reporting on Their Own Evaluation*

- ▶ Although principals and assistant principals reported mixed responses to understanding the evaluation components, in most districts, a majority perceived the leadership practice and student growth measures as fair.
- ▶ In most districts, a majority of principals and assistant principals reported their evaluator as credible, but they did not disagree that their practice rating depended more on their evaluator than their practice.
- ▶ While in 60 percent of the districts, 76 percent or more indicated that the evaluation would lead to better school leadership and improved student learning, they tended to have lower levels of agreement with items pertaining to using feedback to improve **performance, to the evaluation's impact on professional development**, and to making changes to practice.

### *Principal Evaluators*

- ▶ Principal evaluators responded favorably to most survey items.
- ▶ Compared to the other groups, principal evaluators responded more consistently across the survey items.

### *Are Some Districts Doing Better than Others?*

A natural question stemming from our analysis is whether the survey results show that some districts are doing better than others in implementing their evaluation systems. One might expect that in better implementing districts, a higher percentage of educators would have favorable responses to all or most survey items. Preliminary study analyses suggest that for both the teacher and principal evaluation questions, only a few districts

tend to have high proportions of favorable responses on most survey items, and only a few have very low proportions of favorable responses to the survey items. In contrast, many districts tend to have mixed results that do not illustrate a clear pattern. For example, in some districts, there is strong agreement that evaluators and the practice rubric are fair, but less agreement about whether the evaluation results will affect their practice. In other districts, a smaller proportion of respondents agreed that their evaluator and the practice rubric are fair, but higher proportions agree that the evaluation results will change practice. This variability likely stems from the implementation process. Implementing evaluation systems is a difficult task and often takes several years of work. In addition, because districts' implementation plans vary (e.g., some may have implemented some aspects with only a subgroup while others implemented districtwide), it is perhaps not surprising that in most districts, large percentages of educators did not respond favorably to all items. It will be interesting to see whether clearer patterns emerge after another year of piloting or implementation.

# Case Study

In order to gain a deeper understanding of educator perspectives and attitudes toward PERA-compliant evaluation systems, the study team conducted a case study comprising five districts<sup>44</sup> during the 2013-14 SY. The study used a purposive sampling approach, which stratified districts on variables such as region of the state, urbanicity, number of students within district, and SIG versus non-SIG-funded districts, to gather the variation in experiences among participants in the design and implementation of teacher and principal evaluation systems (see Appendix C for the case study methodology and demographics). The case study completed 61 interviews from the following staff roles: principal, teacher, district administrator,<sup>45</sup> union representative, teacher evaluator, and principal evaluator.

Before conducting interviews, the study team analyzed evaluation documents in order to most appropriately focus interview questions. Following the collection of interview data, the team used NVivo qualitative research software to analyze data across and within districts.

This paragraph offers some general findings, followed by the details of the case study. Districts vary in size, urbanicity, and student demographics. Their educator evaluation systems share commonalities and differences. To design and implement the educator evaluation systems, districts developed evaluation committees. The committees represented important stakeholders in the district and were representative of district staff and administrators, teachers' union representatives, teachers, and staff specialists. In most instances, the evaluation committees played an important role in collecting stakeholder feedback during the design process and in disseminating information and supporting peers during implementation. As the evaluation systems move into full implementation, these committees continue to convene and play a role in system refinement.

---

<sup>44</sup> A sixth district agreed to participate in the case study with a request for teacher interviews to be held in August 2014. The study team will submit an addendum to this report to ISBE in late fall 2014, which will incorporate demographics and findings from the remaining district.

<sup>45</sup> Please note that discussions with administrators as a respondent group do not refer solely to principals, but may include school-level administrators such as principals and assistant principals, as well as district-level administrators with a firm understanding of their PERA compliant teacher/principal evaluation systems, such as superintendents and/or human resources administrators. In many instances, participants “wore more than one hat”; for example, a superintendent may have been a principal evaluator, but also a district-level administrator.



Most districts also commonly experienced difficulty linking evaluation results to professional development. For example, some schools within districts had plans for professional development, but these plans were not implemented districtwide. Specific to principal evaluation, the evaluation systems were similar across districts. All districts were implementing each component of the principal evaluation system, including the student growth component for the 2013-14 SY. For teacher evaluation, all districts used similar teacher practice frameworks with embedded observation tools (i.e., Charlotte Danielson's Framework for Teaching). However, districts varied in their plans to communicate, train, and develop buy-in for all staff on the system. Further, some districts (3) had fully implemented student growth models for the 2013-14 SY, while the others (2) did not. Still, those with implemented models experienced issues related to validity.<sup>46</sup>

Below are the detailed results of the case study. The results are organized by theme and in the following order: (1) leveraging evaluation results for professional development, (2) perceiving evaluation system validity, (3) communicating about and training on the evaluation systems, (4) developing educator trust and collaboration on the evaluation systems, and (5) finding adequate time and resources to complete evaluations. Following the results, the conclusion addresses evaluation system early impacts and ongoing implementation.

## **Opportunities to Leverage PERA-Compliant Evaluation for Teacher and Principal Development**

Across all districts, teachers, principals, administrators, and union representatives noted improvements in their current teacher evaluation systems, particularly regarding the opportunity to use observations and professional feedback as occasions to promote educators' development. Interviewees indicated that findings from observations were specific, objective and evidence-based, offering the opportunity for more productive discussion and reflection.

Principals and district administrators valued the inclusion of student growth measures as a part of the principal evaluation systems, but there was a considerable lack of consensus about the utility or value of the observation portion of the principal evaluation system. Many

---

<sup>46</sup> Participants were asked to what extent the evaluation systems, as a whole and their separate components, were valid. The interview questions did not provide participants with a definition of validity, but rather offered several probes that covered the different components that went into an evaluation cycle. Example: How valid do you feel this evaluation system is at measuring your performance (i.e., accurately measures teachers'/principals' true performance)?

- ▶ appropriateness of teacher/principal performance measures
- ▶ domains included in or excluded from evaluation (in-school vs. non-school factors that influence student growth)
- ▶ extent to which evaluation feels objective/tied to evidence vs. subjective/based on opinion
- ▶ capacity of different evaluators to use rating system in the same way.

principals appreciated the opportunity to self-reflect on their professional practice, especially as it related to instructional leadership.

### Teacher Evaluation

While some teachers expressed concerns over the new systems' limited number of observations, others believed the new system incorporated adequate observations, as well as other data points, such as student growth and additional artifacts (i.e., portfolio). The latter teachers felt the new system yielded an ongoing cycle of evaluation over the course of the SY, with multiple data points and methods of data collection, and therefore a fairer process. As one administrator reported,

An added perceived benefit of newly designed PERA-compliant evaluation systems was the opportunity for mid-year feedback and resultant practice enhancements. Several teachers indicated a benefit to receiving results from mid-SY student performance assessments because they could use this information to tailor instruction to students' needs, for example, to provide extra support to students struggling to achieve academic growth. Teachers also perceived that principals and assistant principals visited their classrooms frequently. As a result, these administrators were likely to see teachers implementing best practices and establishing mechanisms to share these best practices as appropriate across grade levels, subjects, or schoolwide. Alternatively, administrators who observed areas of concern during observations also preferred the opportunity to share this information as soon as possible to give teachers the maximum amount of time possible to improve their performance prior to summative evaluation ratings.

### Principal Evaluation

While participants strongly preferred the evidence-based and discussion-oriented nature of the evaluation system, participants still noted challenges specific to principal evaluation. Across the small sample of districts included in this case study, there appeared to be less consistency on the principal evaluation frameworks. From a policy perspective, this will make it more difficult to describe the impact of the evaluation system on principal performance statewide. In addition, some administrators reported that the more varied and abstract nature of principal duties, as compared to teachers, made it difficult to develop an evaluation system that was truly relevant and easy to use. Administrators also saw the large amount of documentation needed to demonstrate principals' professional skills in most districts as unwieldy. A portion of principals expressed the desire for fewer domains within their evaluation, but the opportunity to hone in on specific skills in greater depth. Principals saw this as an opportunity to develop higher level skills as advanced professionals, rather than merely comply with an accountability system. One principal indicated,

### Perceptions of System Validity

Across all districts, teachers, administrators, and evaluators offered varying perceptions on the validity of the newly designed evaluation systems.

### Teacher Evaluation

Teachers expressed more concern about validity than did those in the role of evaluator. All districts that participated in the interviews used the Danielson Framework as the foundation for their teacher evaluation system. Teachers reported satisfaction with the clarity that the Framework offered, regarding the definition of and performance expectations for a high-quality professional educator. Many teachers indicated that they were happy or willing to receive an evaluation based on student growth. Finally, most teachers felt the district was implementing the system as designed.

Teachers' primary concerns about the validity of the system focused on four areas of perceived bias:

1. *The way in which student growth objectives are set or the ways in which student growth is measured.* For example, some districts allowed teachers to set their own targets, while other districts set the target for teachers. Further, some districts permitted teachers to create their own assessments, while others used standardized assessments. In instances where districts used standardized assessments, some were inaccurately calculating growth by measuring growth using different cohorts of students. One teacher spoke to this point as follows,

2. *The timing (during the SY) of student assessments used to determine student growth.* For example, some districts selected assessments that are issued early in the SY in order to get results in time to compute student growth or to include evaluation results in teacher retention/dismissal recommendations, often required by early spring. Teachers were concerned about how much opportunity they have to affect student academic growth by such an early date. One teacher's comment highlights this issue,
  
3. *The frequency and ratio of planned versus unplanned observations.* Many teachers indicated they would like more unplanned observations because they felt anyone can develop a high-level lesson for a planned observation, thus introducing bias in the rating. One teacher spoke of the weaknesses of the current evaluation system this way,

4. *The inconsistency in how evaluators collect and consider additional artifacts to determine a teacher's rating on the two domains of the Danielson Framework that cannot easily be observed (primarily Domains 1 & 4).* One teacher noted,

Principals and administrators noted additional concerns about the validity of the teacher evaluation system. Principals expressed concern that the PERA-required annual timeline does not allow new or less-experienced teachers adequate time to improve their skills between observations. One evaluator noted:

Principals and teacher evaluators also noted that it is challenging to apply the Danielson Framework effectively and demonstrate student growth for staff that serve in specialized instructional roles. In particular, they expressed concerns about the challenges of accurately capturing growth for instructors working with gifted/talented, ELL, or special education students. In these instances, teachers may work with a low number of students on not commonly tested skills, serve highly mobile students, and/or teach students with sustained high performance. Additionally, they noted that the Danielson Framework does not adequately cover other instructional roles with the same level of detail or does not adequately address indicators unique to ELL teachers regarding language acquisition (processes, strategies, assessment of, etc.) Further, some teachers and union representatives cast doubts on evaluators' capacity to have the requisite base knowledge to evaluate and support teachers equally across all content areas. For example, teacher evaluators may not comprehend a foreign language lesson or be familiar with best practices for teaching it.

Teachers, principals, and administrators also noted concerns about a lack of systematic implementation of PERA-compliant teacher evaluation systems. The districts involved in the interviews were at varying stages in the design and implementation process. What was evident from the conversations is that districts often did not implement the systems with a clear plan for collecting feedback formally or informally during initial implementation. In the absence of a formal district plan or process for collecting feedback, principals often felt pulled to react and respond to challenges. Several principals noted that the lack of planned and systematic opportunity for feedback simply creates the condition of “the squeaky wheel getting the grease”—and thus produces questionable system fidelity.

### Principal Evaluation

Noted above, there appeared to be less cross-district consistency in the principal evaluation guiding frameworks. Therefore, consensus varied on perceptions of principal evaluation validity. However, principals and evaluators expressed the following three similar concerns:

1. *The under-developed nature of the indicators used to define principal practice in some of the frameworks.* A number of principals and evaluators commented that the framework used in their district sufficiently defined the scope of principal roles but that the indicators used to define levels of practice were not as well-developed as the teacher framework. For example, one person noted,



2. *The utility and amount of documentation required as a part of the principal evaluation process.* Some principals noted that the framework inaccurately framed their position more as an operations manager rather than an instructional leader. In addition, a number of evaluators and principals noted that the volume of required evaluation documentation increased the likelihood of bias, as some evaluators may refuse to follow all guidelines given their other professional responsibilities. One person noted,
  
3. *The role of evaluation on retention and/or dismissal.* Because many principals are hired at the will of their school board, principals expressed confusion regarding the role of the evaluation. Some principals indicated that even if they demonstrated outstanding performance, the school board could dismiss them at will. One principal stated,

## Ongoing Communication and Training

During evaluation system implementation, teachers, principals, administrators, and union representatives noted the need for consistent ongoing communication, training, and support. At the very beginning, districts trained educators on evaluation system expectations and procedures.

### Teacher Evaluation

Most teachers indicated that they were offered handbooks or reference guides that outlined the same information contained in the training. Some teachers noted that the timing of the training was too separated from the actual implementation. Alternatively, some noted that the training happened concurrently with the final design of the system, and trainers were not fully clear on the new system. Further, some of the districts provided more detailed training that allowed teachers to develop skills on student learning objectives. In the districts where educators expressed more trust and buy-in to the system, the school leadership *“rolled up their sleeves to support our planning and work”* and continued to *“remind us to focus our efforts on student learning and what’s best for kids.”* Across all districts, few teachers knew of ongoing or future training/professional development on the evaluation system or improving professional practice through the system.

All districts implemented the observation portion of their evaluation systems, either districtwide or with a subset of teachers. Through this implementation process, educators reported that they learned that they needed to collaboratively revise the system to improve observation validity and consistency (via recalibration). Principals, administrators, and union leaders noted that recalibration is important to develop and maintain highly trained evaluators, as well as to have an equitable system that teachers support. Teachers expressed concern about evaluator capacity to effectively observe across different grades and subjects. Many teachers expressed hope that as evaluators engage in observation recalibration and the district improves the overall evaluation system, the district will use observation results to direct teachers to professional development that improves their practice.

### Principal Evaluation

The interviews uncovered no notable findings about communication or ongoing training regarding the principal evaluation system.

## Trust/Collaboration

To develop an effective educator evaluation system, educators noted that developing trust, particularly between teachers and administrators, is challenging but very important. At each phase of the evaluation process: design, implementation, feedback, and professional development/remediation, unique trust issues arose.

Many districts used existing structures, such as grade-level meetings and professional learning communities, to share evaluation system information and gather feedback. Smaller districts appeared to have an advantage in sharing information and gathering feedback because they had more direct channels to reach educators. In addition, districts that took advantage of piloting the system to “work out kinks” received more positive feedback and greater trust from educators. Educators were least apt to trust districts with weak communication systems.

Further, district historical decisions and processes may make teachers think that district-level administrators are using the new evaluation system as a tool to dismiss teachers; but, these same teachers may have great respect for and trust their school principal and believe that s/he will fairly administer the system. For example, a trusting relationship between teachers and principals allows for meaningful professional learning. One assistant principal stated,

### Teacher Evaluation

Many teachers expressed the need for confidence in their administrators' ability to use the evaluation system as a mechanism to provide professional development rather than to make retention/dismissal decisions. Districts that reported a mutually trusting relationship between teachers and administrators had an evaluation design committee that included all relevant stakeholders. All stakeholders shared in the development of the evaluation system. One teacher spoke of the collaborative, trusting relationship in his/her district as follows,

Further, by including a diverse group of professionals in the design of the system, the districts were able to more accurately develop a system that took into account all grades, subjects, and student populations.

Conversely, in districts that reported lesser trust between teachers and administration, teachers questioned how administrators would use the evaluation system. Some teachers believed that administrators designed the new evaluation systems to remove teachers and/or demonstrate student gains through standardized testing, rather than for professional development. One teachers' union representative stated, *"This was an accountability measure, a hammer in which to enforce a different way of trying to do business in the district."*

District administrators also seemed to implicitly recognize the greater power they hold in teacher-administrator relationships, given that evaluation ratings contribute in part to teacher dismissal. They frequently indicated that districts were not going to use evaluations as a "gotcha," but as a tool to develop teacher capacity. In addition, among districts that indicated stronger levels of trust, administrators expressed the need to have teachers share in the evaluation development process, as well as to frequently communicate via multiple platforms

evaluation system development. One teacher spoke of the way his/her principal garnered trust as follows,

### Principal Evaluation

Contrary to variable views of teachers regarding trusting relationships with their supervisors, principals reported trusting their supervisors (and evaluators). Although some principals felt the evaluation practice standards could be overly vague and/or too numerous, they indicated that the evaluation system was fair, rooted in evidence, and captured their overall duties. Principal evaluation more heavily weighted student performance data, but many principals indicated that this was reasonable and felt that in their role as an instructional and operational leader, they should be accountable for student performance in their school. One principal evaluator spoke to the strengths of his/her principal evaluation system as follows,

## **Time/Effort/Skill to Complete Evaluation Cycle**

Across all districts, teachers, principals, district administrators, and union representatives expressed their concerns regarding the lack of time to design and prepare for the evaluation system rollout. As a result of time constraints, educators reported that introductory training was insufficient to build a true understanding of how the new evaluation systems differ from prior systems. In addition, as a result of limited time, educators noticed that evaluators lacked comprehensive knowledge of the systems. While most evaluators completed the state-provided training on how to conduct observations and how PERA compliant evaluation systems operate, many educators felt evaluators lacked necessary training in how to best engage staff in coaching conversations as part of post-observation conferences and summative feedback, as well as how best to translate findings to actionable next steps for professional development. One teacher reported,

Due to lack of time, many districts noted that their pilot systems were unable to fully work out design and implementation concerns. Concerns included how to apply observation rubrics to a range of professional roles, how best to develop goals for student academic growth and collect supporting data, and how to address logistical concerns, such as challenges in the technology that supported districts' evaluation systems. These unresolved concerns prompted some teachers to question how districts could effectively provide performance ratings, particularly in light of the attached consequences, such as staff remediation and retention decisions. One teacher stated his/her opinion as follows,

Districts that were able to accommodate a more successful pilot year or a year where student growth data did not tie to **teachers'** performance rating indicated higher trust and satisfaction among staff members.

### Teacher Evaluation

Educators magnified the significant effort required to complete evaluation-related tasks, such as conducting observations, preparing and reviewing artifacts of educator performance, and analyzing student growth data. Time burdens fell most heavily on a limited number of administrators responsible for carrying out a high volume of teacher and staff observations. Considering that school-level administrators could be responsible for evaluating as many as 100 staff members per year, the formal observation process alone was estimated to take up to 1,200 hours of administrator time per SY, not including other evaluation components, such as informal observations, student growth measurement, and artifact review. Formal observations became a task that was competing with administrators' other professional duties. As one teacher indicated,

In part, this time burden appeared to be exacerbated by the limited number of roles within a school or district for those who are both qualified to conduct evaluations and permitted by law or union contract to do so. For example, teachers, union representatives, and administrators in some districts indicated that teachers were specifically barred by contract from observing each other for the purposes of performance evaluation. Smaller and/or less resourced districts were challenged by the limited number of professionals available who legally were able to observe principals in their district. As one administrator pointed out,

## Principal Evaluation

Specific to the design of principal evaluation systems, districts supported using a criterion and evidence-based system of evaluation that emphasizes the instructional leadership portion of the principal's role. Overall, principals, administrators, and teachers accepted student growth data as a portion of principals' evaluation, given their role as an educational leader for the entire school. However, districts expressed some challenges in observing a principal's performance due to the more abstract and varied nature of his/her role. Principal evaluators, administrators, and principals also noted that there were a great number of domains to observe for principals as well as a high volume of artifacts to submit as evidence of their performance. As one principal stated,

## Early Impacts

Drawing on the findings above, there are several early evaluation system impacts to highlight:

1. The teacher evaluation process fosters dialogue focused on improving (1) teaching practice and (2) student learning.
2. The teacher evaluation system uses a transparent, criterion-based framework that increases clarity, consistency, and objectivity in evaluating teachers and provides a platform for more useful coaching conversations.
3. The teacher and principal evaluation system helps principals focus on fostering high-quality building instruction and student academic growth.
4. The principal evaluation system enhances dialogue between principals and district leaders and develops principals as instructional leaders.



## Enhancing Ongoing Implementation

As districts and policymakers continue to enhance the evaluation systems, the case study findings shed light on three key issues that help districts implement and sustain evaluation systems.

1. *Ensuring trust:* A growing body of case study and clinical narratives<sup>47</sup> note how relational trust is required for meaningful school improvement. At the district or school level, this trust includes collective educator decisionmaking with broad teacher buy-in, a natural diffusion of reform initiatives, a desire to create district or school improvement, and more timely and effective implementation of reforms. The districts participating in this case study demonstrated a range of instances where relational trust facilitated more timely and effective evaluation system design and implementation. Similarly, districts with less trust faced barriers to effective evaluation system design and implementation.
2. *Aligning evaluation and professional development:* Though it may very well be a condition of the developmental stage of the evaluation systems, districts struggled to concretely articulate a plan for linking current and future educator professional development opportunities to evaluation results. Although some districts have adopted TeachScape and its related professional development modules, there still remains a gap in linking comprehensive professional development to individual evaluation results. Some evaluators spoke to the misalignment between professional development and evaluation results and indicated a desire for more resources to support teachers in need of development.
3. *Sustaining the new evaluation systems:* The design and implementation of a PERA-compliant evaluation system requires considerable investment in human resources in the districts across the state. In a time of strained budgets and greater accountability demands, it appeared that there are concerns about the demands to sustain the evaluation systems. As districts face staff turnover in the coming years, there will be ongoing needs for training teachers, principals, and evaluators to effectively use the systems. The demands on the time of the principal to conduct a complete evaluation with each teacher makes it considerably more challenging for a principal to devote adequate time to roles other than instructional leadership. Some districts are exploring ways to address this time and effort drain. For example, it may be possible to use external observers or peer evaluators to conduct a portion of educator observations.

---

<sup>47</sup> Bryk, A., and Schneider, B. (2002). *Trust in schools: A core resource for improvement*. New York: Russell Sage Foundation.



# Conclusions

*This chapter reviews the major conclusions drawn from the document review, survey, and case study findings. Using the conclusions as a basis, the research team makes several recommendations to enhance the PERA educator evaluation systems. The organization of this chapter is as follows: (1) a brief summary of answers to the research questions, (2) a discussion on important additional findings, and (3) an overview of recommendations.*

## Research Questions

Indicated in the beginning of the interim report, the team used two primary questions to frame this phase of the study.

- ▶ What did districts do in SY 2013-14 to implement PERA-compliant teacher and principal performance evaluation systems?
- ▶ What are the perceptions of teachers, principals, and the evaluators about the evaluation systems?

Each primary question includes additional underlying questions.

This mixed-methods study provided multiple lenses for the research team to explore the **primary and underlying questions**. The section below offers a synopsis of the study's findings for the primary and underlying questions.

### Question 1: How did districts evaluate teacher and principal professional practice?

Participating study districts' teacher evaluation systems included a professional practice framework that follows PERA requirements, including: (1) the use of rubrics with specified rating levels, (2) the inclusion of the appropriate number of observations, and (3) the training of evaluators. The majority of districts used the Danielson Framework for Teaching to evaluate teacher professional practice.

For principals and assistant principals, most districts used a professional practice framework that included: (1) conducting two formal observations of principals and assistant principals; (2) informing principals and assistant principals about evaluation requirements; (3) having principals, assistant principals, and their evaluators set principal and assistant principal performance goals; (4) requiring principals and assistant principals to conduct self-evaluations; (5) providing evaluation feedback to principals and assistant principals; and (6) offering mandatory evaluator training. Further, the majority of districts used the Illinois Performance Standards for School Leaders rubric; although some used significantly different frameworks.

#### *1b. How did districts measure student growth?*

Of the districts (28) that provided documents on their teacher evaluation systems, 24 piloted growth scores for teachers during the 2013-14 SY. Twenty two of these districts used Type I and II assessments, while 15 used both Type I/II and Type III assessments, and 2 used only Type III assessments. The Type I and II assessments, including the state tests (ISAT and PSEA/EPAS), proprietary assessments, externally developed and scored assessments, and district-developed common assessments, varied across districts.

Thirty of the 32 districts that provided documents on principal evaluation reported using student growth as a part of their principal and assistant principal evaluation systems and combined student growth measures with practice and other measures to determine an overall evaluation rating for the 2013-14 SY. Districts used a variety of Type I and II assessments, including ISAT or PSEA/EPAS,<sup>48</sup> externally developed and scored assessments, and district-developed common assessments to calculate a student growth score.

---

<sup>48</sup> Typically, these districts used one-year lagged assessment data and defined student growth as an increase in the percentage proficient score level between two years.

To measure growth for teachers and principals/assistant principals, most districts used or intend to use growth targets or goals on Type I, II, or III assessments. These districts based the student growth rating on either the number of goals met or exceeded or an average of ratings made on each goal (i.e., teachers received a 4 if growth was above the target, 3 if it met the target, etc.). Only two districts used or intended to use more sophisticated value-added models for measuring growth on Type I or II assessments for teachers, and only three districts used or intended to use these models for measuring student growth for principal and assistant principal evaluations.

***1c. How did districts combine ratings of teacher and principal professional practice with student growth to determine an overall summative evaluation rating?***

For the 2013-14 SY, 11 districts gave teachers a summative evaluation score that combined growth with practice. Of these 11 districts, 5 used a decision matrix to combine the scores, while 6 used a weighted compensatory method. The districts using a weighted compensatory method weighted practice at either 70 or 75 percent and growth at 25 or 30 percent. The remaining study district had either not yet piloted a combined rating or did not provide sufficient information to make a determination about its method for combining measures.

For principals, all but two of the 32 districts that shared documents had fully operational principal evaluation systems that measured both leadership practice and student growth in the 2013-14 SY. Most (29) of these districts used a weighted compensatory method to combine growth (including other student outcomes) and practice ratings for principals. Districts used various weighting schemes ranging from 75 percent on practice and 25 percent on growth to 50 percent on practice and 50 percent on growth.

***1d. How did districts communicate about their evaluation systems to teachers and principals?***

Almost all the districts that shared evaluation information developed a teacher evaluation handbook or similar document. However, many of these districts had not fully documented the student growth component of the process (in some cases because that component had yet to be finalized at the time of the document collection). Despite the absence of information on student growth, almost all of these districts provided some form of training on the new evaluation system to teachers, varying from one-on-one orientations with evaluators during a pre-conference to providing 40 hours of training on the rubrics and processes alongside the evaluators. In addition to the information provided to teachers, all districts that shared documents appeared to have informed principals of performance expectations and criteria for evaluation, though the amount of information districts provided varied. A few districts

developed a principal evaluation guidebook that described both the practice and growth components in detail, but most districts provided resources with relatively little information regarding their strategy for measuring student growth.

Teacher, principal, and assistant principal survey results are consistent with the findings from the document review. The teacher survey results indicated that in the vast majority of districts, teachers received information on the evaluation system from multiple sources, ranging from websites to one-on-one meetings with evaluators. In addition, in the majority of districts, most teachers rated these sources of information as useful or very useful. The principal survey results indicated that principals and assistant principals in the majority of districts reported that they had received notification about the criteria for evaluation.

The results from the case study districts also align with the finding that districts provided information on the teacher and principal evaluation systems. In the case study districts, most teachers indicated they were offered handbooks or reference guides covering the evaluation process and that local joint committees were an important factor in communicating about the evaluation process during the design and implementation of the evaluation process. In particular, especially in the smaller districts, it appeared committees were an important channel of communication to those being evaluated.

*1e. In what areas have districts made progress toward full implementation of PERA-compliant teacher and principal evaluation systems, and in what areas are there concerns or difficulties?*

Districts have made substantial progress implementing the professional practice evaluation provisions of PERA. The document review suggested that all or most of the districts that provided documents designed systems that follow PERA requirements on the practice rubrics, the number of observations, and the number of rating levels. In addition, most districts provided information on the evaluation process, trained evaluators, and took actions to promote inter-rater agreement. Notably, however, it was less common for districts to promote inter-rater agreement or consistency for evaluators of principals.

The case study and survey results confirm evidence collected from the document review that districts are complying with PERA requirements on evaluation of professional practice. In the case study districts, teachers perceived that the systems were being implemented as designed. Survey results showed that, in the vast majority of districts, teachers received information on the evaluation system; non-tenured teachers received the required two observations; and teachers had a good understanding of the practice rubric or framework. Similarly, in the majority of districts, principals and assistant principals reported they had received rubrics,

met with evaluators to set goals, completed a self-assessment, and received two observations or school visits. It is worth noting that all principals reported receiving feedback within 10 days of a school observation or visit in fewer than half of the districts.

Study results suggest that districts could improve inter-rater agreement processes. The document review indicated that most districts, while providing resources to ensure inter-rater agreement, did not actually follow up to check rater agreement. Moreover, it appeared that few districts provided resources on ensuring inter-rater agreement among principal evaluators (though in some cases there were only two evaluators in the district). In addition, the survey and case study results confirm the need for more emphasis on rater agreement. For example, while teachers who had received evaluations and responded to the survey appeared to have confidence in their evaluators, it was unexpected that in over 90 percent of the districts, 51 percent or more of both tenured and non-tenured teachers agreed that ratings depended on the evaluator. In addition, though in most districts, 51 percent or more of principals and assistants reported being moderately or very confident that their ratings would be consistent with other raters, only in 15 percent of the districts were these respondents very confident. In the case study districts, principals, administrators, and union leaders noted that recalibration of raters is important to maintain evaluator skills and to show teachers that the ratings will be equitable. These results suggest the need to continue to provide opportunities for evaluators to discuss how evidence and rubrics are interpreted, assess inter-rater agreement, and communicate those results to teachers and school administrators to ensure them of the consistency of interpretation across evaluators.

Though districts have made substantial progress implementing the student growth measures for principal and assistant principal evaluations, they have made less progress on measuring student growth for teacher evaluations. Almost all the districts that submitted documentation developed procedures and are using student growth as part of their principal evaluation systems. Furthermore, the districts have developed PERA-compliant methods for combining practice and growth ratings to derive an aggregate evaluation rating. However, several districts that submitted documents had not yet fully piloted the student growth component of the teacher evaluation process and did not appear to have documented the process for measuring student growth as well as they had the practice component. This issue arose in the survey data as well when teachers, principals, and assistant principals reported that their understanding of the student growth measures and the method of combining practice and growth ratings was weaker than their understanding of the practice evaluation.

Study results also suggest that a substantial number of districts still need to work on a system for ensuring accurate teacher-student links, establishing guidelines for minimal numbers of students to include in growth measures, handling student mobility, and systematically accounting for student characteristics. At the time of the document review, only the districts

using value-added systems had a systematic method for assessing the precision or potential reliability of their growth measures.

Additionally, in many districts, it was difficult to understand the criteria for how student growth translated into a growth rating and whether growth goals were set in a way that would allow a fair comparison across teachers, principals, or assistant principals within the district. At least one case study district appeared to calculate growth for teacher evaluations as the difference in assessment scores between two different cohorts of students, which is not consistent with the spirit of PERA. It also appeared that some other districts are doing this for principal evaluation in place of more refined growth measures.

### Question 2: What are the perceptions of teachers, principals, and the evaluators about the evaluation systems?

Many researchers studying performance evaluation agree that the perceptions of those evaluated and their evaluators are important for the long-term sustainability and effectiveness of evaluation systems.<sup>49</sup> In particular, both groups need to understand the system and believe it has a positive impact on performance. Those being evaluated should also perceive the system as fair, their evaluators credible, and that they receive useful feedback. The surveys, as well as case study protocols, asked teachers, principals, assistant principals, and principal evaluators about these topics.

#### *2a. Did teachers, principals, and evaluators understand the evaluation systems that districts piloted or used in the 2013-14 SY?*

Evidence from the surveys indicates that in most districts, a majority of teachers, principals, and evaluators had a strong level of understanding of the professional practice rubrics, but not as strong an understanding of the student growth measures or how their districts combine the measures to determine a final evaluation rating. For example, in the average district, 60 percent of the teachers rated their understanding of the rubric as strong or very strong, but only 49 percent and 48 percent rated their understanding of the student growth measures and the method of combining measures as strong or very strong. For principals and assistant principals, 75 percent rated their understanding of the practice rubric as strong or very strong, while 47

---

<sup>49</sup> For example, DeNisi, A.S., and Pritchard, R.D. (2006). Performance appraisal, performance management, and improving individual performance: A motivational framework. *Management and Organization Review*, 2:2, 253-277; Levy, P.E., and Williams, J.R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, 30, 881-905; Heneman, H.G. III, and Milanowski, A.T. (2003). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education*, 17:2, 173-195.



percent and 51 percent rated their understanding of the student growth measure and the district's strategy for combining measures respectively as strong or very strong.

**2b. *Did teachers and principals perceive these systems as fair?***

Teacher and principal survey responses by district indicate that, on average, they perceived the professional practice evaluation as fair, but had less favorable perceptions about the fairness of the growth measures. According to the teacher survey, in the average district, 78 percent of the responding teachers agreed or strongly agreed that the practice rubric was fair, while only 51 percent agreed or strongly agreed that the student growth measures based on Type I or II assessments were fair. There was slightly more support, with 72 percent agreeing or strongly agreeing that the measure was fair, for student growth measures based on Type III assessments. In addition, the case study results support this finding by reporting that teachers felt satisfaction with the clarity of the Framework for Teaching and its definition of high-quality teaching. The principal survey responses were similar. In the average district, 88 percent of principals and assistant principals responding agreed or strongly agreed that the leadership practice part of the evaluation was fair, while 66 percent agreed or strongly agreed that the student achievement growth measures were fair.

**Case study districts offer detail about teachers' and principals' concerns with the evaluation systems in their districts.** Teachers noted several concerns about the validity of the systems. They mentioned that there was often variability across teachers in how growth goals are set and growth is measured. They also reported that the requirement that they complete evaluations in March gives teachers less time to affect their student growth measures. With respect to the evaluation of practice, they reported that unplanned observations may be more representative of practice than one or two planned observations. In addition, they believed that there was inconsistency in how evaluators collect and consider additional artifacts to determine ratings on the two Danielson framework domains that cannot easily be observed. Principals largely indicated that the principal evaluation system was fair and that it was fair to receive an evaluation based on student growth. However, principals and their evaluators expressed concerns that the indicators used to define levels of practice were not as well developed as the teacher framework, making it difficult for evaluators to apply them in a consistent way.

### *2c. Did teachers, principals, and evaluators perceive that the evaluations affected teacher and principal professional development?*

In most districts, teachers, principals, and assistant principals agreed that they received feedback that was potentially useful to improve their practice. On the teacher survey, the majority of teachers in every district agreed or strongly agreed that the feedback they received identified specific areas for improvement and included guidance or suggestions for making improvements. They subsequently used the feedback to improve instruction. In most districts, the majority of principals and assistant principals responding to the survey agreed or strongly agreed that the feedback they received identified specific areas that could be improved and included guidance on how to make improvements. However, there was less agreement about whether feedback included suggestions for specific professional development.

In most districts, a majority of teachers, principals, and evaluators responding to the surveys perceived that the evaluations influenced teacher or principal professional development. In most study districts, large majorities of teachers responding to the survey agreed that the observations encouraged them to reflect on their practice and that their observation rating would influence their professional development activities. In contrast, however, smaller majorities of teachers in most districts agreed that the student growth measures would inform future professional development. In addition, in the majority of districts, principals and assistant principals who evaluated teachers agreed that the evaluation process had or would **affect teachers' professional development choices and that they used evaluation results to focus school professional development activities.**

Similarly to teachers, in most districts, the majority of principals and assistant principals responding to the survey items about their own evaluations agreed that the evaluation process encouraged them to reflect on their practice and influenced their professional development activities. Further, like teacher evaluators, in most districts, the majority of those evaluating **principals agreed that the evaluation process influenced principals' professional development choices and helped focus professional development resources where most needed.** The case study results confirm these findings.

In the case study districts, participants mentioned that the new evaluation process encourages dialog on improving teaching practice and student learning. Another positive impact noted was a stronger focus of principal evaluation on instructional leadership. In addition, in every case study district, teachers, principals, district administrators, and union officials noted that the new teacher evaluation systems provided better opportunities to use the observations and feedback to promote teacher development.

*2d. Did teachers, principals, and evaluators perceive that the evaluations affected instruction or leadership practice?*

In most districts, a majority of teachers, principals, and evaluators indicated that teachers or principals made changes to their instructional or leadership practices as a result of the evaluation process, but principal evaluators were more optimistic about changes in principals' practices than teacher evaluators were about changes in teachers' practices. Further, most districts' tenured and non-tenured teachers who responded to the surveys agreed that they made changes to their teaching as a result of the observation process. Additionally, in the majority of districts, principals and assistant principals who evaluated teachers agreed that teachers had changed the way they taught in order to do better on the evaluation. In the average district, the majority of these school leaders responded that about half of the teachers they evaluated incorporated their feedback and made noticeable improvements in teaching.

Principals' and principal evaluators' responses were more similar. In most districts, the majority of principals and assistant principals responding about their own evaluation agreed that they had made changes in their leadership practice as a result of the evaluation and that they changed what they do in order to improve their student growth measures. Evaluators of principals and assistant principals responding to the survey largely agreed. In over 75 percent of the districts, the majority of these evaluators agreed that principals changed their behavior to do better on both components of the evaluation process.

## **Additional Findings**

There is evidence of some early positive effects of PERA.

While it is too early to see a strong impact of the new evaluation systems on student learning,<sup>50</sup> results from the surveys suggest that the initial use of PERA-compliant evaluation systems has had some positive impacts. First, the process appeared to have influenced educator professional development. As discussed above, on the surveys teachers, principals, and their evaluators perceived that the evaluation process influenced staff development for both teachers and principals. Second, teachers, principals, and evaluators thought that the process affected instruction and leadership practice. In most districts, a majority of teachers, principals, and evaluators responding to the surveys indicated that teachers or principals made changes to their instructional or leadership practices as a result of the evaluation process.

---

<sup>50</sup> Note that all of the Race to the Top districts except Chicago Public Schools were piloting PERA-compliant teacher evaluation, and this was only the second year of PERA-compliant principal and assistant principal evaluation.

Teachers, principals, and evaluators of principals are optimistic that PERA-compliant evaluation systems will have positive impacts in the future.

- ▶ In the average district, 66 percent of teachers responding to the survey agreed or strongly agreed that the evaluation process would lead to better instruction and 61 percent agreed it would lead to better student learning. The corresponding percentages for principals and assistant principals who were asked about teacher evaluation were 93 percent and 82 percent, respectively.
- ▶ In the average district, over 90 percent of the evaluators of principals responding to the survey agreed that the evaluation process would lead to better school leadership and better student learning. In the average district, over 75 percent of the principals and assistants reporting about their own evaluation agreed or strongly agreed.

The time and effort needed to carry out PERA's requirements for more rigorous evaluation is likely to burden some districts and is a concern of those receiving evaluations, and to a lesser extent, of the evaluators.

- ▶ The document review revealed that a substantial minority of districts did not have the staff resources to quickly implement PERA-compliant evaluation systems and that this was one reason why some had not fully piloted the student growth component.
- ▶ In all the case study districts, participants mentioned a lack of sufficient time to design and roll out the PERA-compliant evaluation systems. There were also concerns about sustaining the burden required by these evaluation systems, including training new evaluators as staff turns over and principals finding the time to balance their duties as an evaluator with their other responsibilities.
- ▶ Educators in the case study districts spoke of the substantial effort required of evaluators to carry out the evaluation process. The time required for multiple classroom observations of teachers was a concern. In smaller districts, time demands were exacerbated by the limited number of qualified evaluators in the district. The amount of documentation needed to substantiate principal performance ratings was also seen as unwieldy.

- ▶ The survey questions asked evaluators of teachers and principals to estimate the time it takes to evaluate one teacher or principal. In the median district, evaluators reported the average time to evaluate each teacher was 5 hours<sup>51</sup> and each principal was 11 hours. While these averages do not seem unusually high, they likely represent more time than evaluators spent before PERA.
- ▶ In most districts, the majority of teachers and principals tended to agree that, with respect to their own evaluations, the process takes more time or effort than the results are worth.
- ▶ In their role as evaluators, principals and assistant principals generally did not agree that teacher evaluation takes more effort than the results were worth. The majority of these respondents agreed in only 20 percent of the districts. The majority of principal evaluators agreed only in 16 percent of the districts.
- ▶ Time demands were the dominant theme in comments made by principal evaluators and principals who evaluated teachers in response to the open-ended questions that allowed for additional comments on the evaluation process.

To get the most out of the potential of the new evaluation systems to improve staff development, some districts may have to invest more effort to connect evaluation to professional development.

- ▶ As mentioned above, during the case study interviews, educators indicated that the new systems had the potential to improve teaching. However, the case study results also suggested that districts can find it difficult to link evaluation results to actual professional development activities. None of the case study districts articulated a plan for linking current or future professional development opportunities to evaluation results.
- ▶ In only 39 percent of the districts did the majority of principals and/or assistant principals report that the feedback they received included suggestions for specific professional development that could help improve their performance.

---

<sup>51</sup> This estimate does not include time to calculate growth measures and combine growth and practice ratings because this is likely to have been done at the central office level.

Flexibility provided by PERA resulted in the adaptation of evaluation systems to local conditions and needs. But this local adaptation could reduce the comparability of ratings across districts.

- ▶ In both the case study and the document review, it was apparent that districts varied in the way they set growth goals, measured growth, and translated growth into a final growth rating. There were also a considerable number of districts that used frameworks other than the Illinois Performance Standards for School Leaders rubric for principal practice evaluation. While the ability of districts to choose their own methods likely helped promote local acceptance of PERA-compliant evaluation systems, it does make it harder to compare results across districts.
- ▶ Further, the cases study and document review also showed that in some districts teachers and school administrators could choose their own assessments and set their own student growth goals, even for Type I and II measures. While this can allow focusing on areas that need improvement, it can also lead to lenient ratings and make comparisons within districts less valid.

# Recommendations

*The study's findings and conclusions suggest seven recommendations that ISBE and PEAC may want to consider. The seven recommendations are detailed below and focus on the following issues: student growth measurement (recommendations 1 – 4), inter-rater agreement or consistency (recommendation 5), principal engagement (recommendation 6), and professional development (recommendation 7). Following the seven recommendations, the section concludes with a brief discussion on PERA program monitoring.*

1. It may prove beneficial to provide districts with more intensive technical assistance on developing student growth measures. Despite the very useful ISBE and PEAC guidance documents on growth measures (notably student learning objectives), many districts need additional and/or more directive guidance. For example, ISBE and PEAC may want to consider developing a more comprehensive dissemination process for materials and actively follow up with districts to ensure their understanding and use. In addition, since many districts are setting achievement goals using Type I and II assessments that resemble SLOs, ISBE and PEAC may want to create technical assistance materials that are generalizable beyond SLOs that focus on setting growth goals, measuring their attainment, and converting goal attainment into growth ratings.
2. Districts may benefit from more specific models for measuring student growth, and methods for addressing data and measurement quality. In addition, ISBE may want to provide greater technical assistance on these issues by leveraging the support of regional universities or educational agencies. As discussed above, many districts have not considered or developed rules to account for the full range of measurement quality issues related to ensuring the validity of student growth measures, such as: minimum numbers of students, student-teacher linkages, student mobility, and student demographic characteristics.

3. Developing more prescriptive guidance on measuring student growth using Types I and II assessments to facilitate comparisons that are more valid across teachers may prove useful. Using two student growth measures based on two different assessments (Types I/II and Type III) allows for both a personalized growth measure that teachers can participate in setting (based on the Type III assessment) and a more uniform and comparable measure (based on the Types I/II assessment). By developing measures based on Types I/II assessments that resemble those developed for Type III assessments, the advantage of using two different assessments is reduced. To maximize the benefit of these different assessments, the study team suggests one of the options below.
  - a. Require districts to use more standardized and uniform growth measurement models, such as value-added, student growth percentiles, or similar statistically motivated models for Types I/II assessments. These methods allow for better estimates of the reliability of the growth measures and can systematically account for student characteristics that affect achievement that are beyond teacher or school control. This option likely requires more support (via technical assistance or regional consortia) for smaller districts to develop and implement these models.
  - b. Develop a state-administered value-added or similar model that compares growth across the state. Such a system maximizes economies of scale and provides measures with a higher and uniform level of validity than many districts could achieve on their own. Districts could then be given the option of using results from this model in their teacher (and potentially also principal) evaluation systems instead of developing their own statistical model.
  - c. Develop a state-administered value-added or similar model for districts to use as a yardstick to assess the validity and rigor of local student growth measures. The state-developed model would not be used to produce the measures for evaluations, but rather as a yardstick to compare with district growth measures which would be used for evaluation. If districts develop local well-constructed measures and apply them systematically, the state and district model should correlate substantially. In districts with low correlations, the state can provide technical assistance to improve the measures.



4. Revisit and clarify the distinction between growth measures based on Type I and Type II assessments. Noted in the document review chapter, despite definitions and examples in the PERA guidance, many districts are unclear about the differences between the assessments and their respective measures. Further, as mentioned above, many districts use very similar measurement models for Types I/II and Type III assessments. ISBE and PEAC may want to consider the role and goal of these distinctions. To ensure that districts create different measures based on the assessments and clarify the existing misunderstandings, ISBE and PEAC may want to define two or three types of growth measurement models (rather than types of assessments) and list the types of tests that can be used for each. If the state retains the distinction between Type I and II assessments, it should consider disseminating clarifying documents and providing technical assistance.
5. Provide more guidance and technical assistance on achieving inter-rater agreement or consistency among teacher and principal/assistant principal raters. Suggestions include: (1) documenting and sharing district best practices, (2) describing and disseminating model rater calibration and re-certification programs,<sup>52</sup> and (3) encouraging districts to do co-observations (where an educator is observed by two evaluators who each make independent ratings) and examine the rater agreement for a sample of those evaluated. Smaller districts with fewer evaluators that use similar rubrics could be encouraged to join in local consortia to hold regular calibration sessions and facilitate co-observations. Districts should then be encouraged to share information on these programs, including the results, with those evaluated. This sharing process will help educators have more confidence in the system.
6. Further develop the Illinois Performance Standards for School Leaders rubric and find ways to involve principals more closely in this process. Some districts adopted the state rubric and process without involving stakeholders (i.e., principals) to the same extent as they did for the teacher process. In addition, the state rubrics might be streamlined to require less work to apply and clarify some of the standards. Further, ISBE and PEAC may want to provide districts with additional guidance on how to adapt the rubrics to the specific job duties of assistant principals.

---

<sup>52</sup> For example, in order to help evaluators stay in calibration, some districts have evaluators meet periodically to view and rate video clips, or discuss how they are interpreting the rubrics in the field. Some require evaluators to pass a recertification test after a substantial interval from the initial training, to ensure that evaluators have not drifted apart in how they are applying the rubrics.

7. To maximize the benefit of the rigorous PERA performance evaluation processes, districts should be encouraged to develop a stronger link between evaluation results and professional development. Study findings show that educators perceive that the evaluation processes have the potential to influence professional development activities. In order to ensure this linkage, districts should systematically evaluate their professional development programs to ensure that opportunities directly related to the evaluation rubric performance dimensions are available and communicated to educators. For example, several federal TIF grantees have developed web-based professional development that links specifically to evaluation rubrics so that teachers can access resources and view video examples of instruction that illustrates higher levels of performance. This process also ensures that the new evaluation system aligns tightly with district improvement efforts and integrates with other components of district human resource management strategies. ISBE and PEAC may want to consider recommending that statewide or regional consortia be formed in order for districts to pool their resources and develop similar web-based professional development. ISBE and PEAC may also want to develop a protocol for districts to use to assess the relevance and coverage of existing professional development programs to the performance evaluation dimensions.

Last, in addition to the recommendations above, the study team suggests that ISBE build on its existing monitoring processes to develop a systematic approach that allows the agency to even more effectively monitor the statewide rollout of the PERA-compliant evaluation systems. Our team recommends the development or refinement of two specific strategies: (1) a district annual submission of a performance report that uniformly documents district implementation and ongoing evaluation system progress and (2) an annual or biannual state onsite or desk compliance monitoring that uses a specific protocol (see for example, appendix D) to guide conversation and systematically collect data. Both monitoring techniques will enable the state to uniformly collect important data from all districts to evaluate individual district and statewide progress and also implement important technical assistance outreach strategies. As the study team continues to work with ISBE to evaluate PERA progress, it will work the agency to discuss how these monitoring strategies may be tailored to best meet their needs to ensure PERA evaluation system success.

# References

Aimsweb (n.d). *Aimsweb products and services*. September 9, 2014. <http://www.aimsweb.com/assessments>

Bill & Melinda Gates Foundation (2013). Ensuring fair and reliable measures of effective teaching: **Culminating findings from the MET project's three-year study**. Seattle, WA: Author.

Bryk, A., and Schneider, B. (2002). *Trust in schools: A core resource for improvement*. New York: Russell Sage Foundation.

California Department of Education (2009). *Standardized testing and reporting*. September 9, 2014. <http://www.starsamplequestions.org/>

CEC Partnership Groups (n.d.). *Growth Through Learning. Illinois performance evaluation*. September 9, 2014. <http://www.growththroughlearningillinois.org/>

The Danielson Group (2013). *The Framework for Teaching*. September 9, 2014. <http://danielsongroup.org/framework/>

DeNisi, A.S., and Pritchard, R.D. (2006). Performance appraisal, performance management, and improving individual performance: A motivational framework. *Management and Organization Review*, 2(2), 253-277.

Dynamic Measurement Group (n.d.). *What are DIBELS?* September 9, 2014. <https://dibels.org/dibels.html>

Discovery Education (2008). *Discovery education assessment*. [http://www.discoveryeducation.com/pdf/assessment/Discovery\\_Education\\_Assessment\\_Research.pdf](http://www.discoveryeducation.com/pdf/assessment/Discovery_Education_Assessment_Research.pdf)

Heneman, H.G. III, and Milanowski, A.T. (2003). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education*, 17(2), 173-195.

Illinois General Assembly (n.d.). *Full text of sb0315*. September 9, 2014. <http://www.ilga.gov/legislation/fulltext.asp?DocName=09600SB0315enr&GA=96&SessionId=76&DocTypeId=SB&LegID=&DocNum=0315&GAID=10&Session=>

- Illinois State Board of Education (n.d.). *Administrative code, part 50, sub a, sec. 50.30*.  
<http://www.isbe.net/rules/archive/pdfs/50ARK.pdf> Illinois
- Illinois Association of School Administrators (n.d.). *Illinois performance standards for school leaders rubric evaluating practice of principals*. September 9, 2014.  
<http://illinoisasa.wikispaces.com/file/view/IL+Principal+Evaluation+State+Model+Summary+with+Appendices+21512.4.pdf>
- Illinois State Board of Education (n.d.). *Illinois standards achievement test (ISAT)*. September 9, 2014. <http://www.isbe.net/assessment/isat.htm>
- Illinois State Board of Education (n.d.). *Performance evaluation advisory council state model for principal evaluation*. [http://www.isbe.state.il.us/peac/word/peac\\_prin\\_eval\\_model.pdf](http://www.isbe.state.il.us/peac/word/peac_prin_eval_model.pdf)
- Illinois State Board of Education (n.d.). *Prairie state achievement examination (PSAE)*. September 9, 2014. <http://www.isbe.net/assessment/psae.htm>
- Illinois State Board of Education (2013). *Guidance document 13-09 model teacher evaluation system creating a summative rating*. September 9, 2014.  
<http://www.isbe.state.il.us/peac/pdf/guidance/13-9-te-model-summ-rating.pdf>
- Illinois State Board of Education (2013). *Guidance document 13-11 guidance on creating a summative rating in teacher evaluation systems*. September 9, 2014.  
<http://www.isbe.state.il.us/peac/pdf/guidance/13-11-te-summ-rating.pdf>
- Illinois State Board of Education (2014). *Joint committee guidebook: implementing the student growth component in teacher and principal evaluation systems*. September 9, 2014.  
<http://www.isbe.net/PEAC/pdf/student-growth-component-guidebook.pdf>
- iSteep, LLC (n.d.). *System to enhance education performance*. September 9, 2014.  
<http://www.isteep.com/login.aspx>
- Levy, P.E., and Williams, J.R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, 30, 881-905
- NWEA (2014). Northwest Evaluation Association. September 9, 2014. <https://www.nwea.org/>
- Rowan, B., Schilling, S.C., Spain, A., Bhandari, P., Berger, D., and Graves, J. (2013, December). *Promoting high quality teacher evaluations in Michigan: Lessons from a pilot of educator effectiveness tools*. Ann Arbor: University of Michigan Institute for Social Research.
- STI (2013). *Education data management solutions*. September 9, 2014. <http://www.sti-k12.com/stias/>
- Student Island (2013). *Study island*. September 9, 2014. <http://www.studyisland.com/>
- Val-Ed (n.d.). *Vanderbilt assessment of leadership in education*. September 9, 2014.  
<http://www.valed.com/>

# Document Review Protocol

## *Part 1: Teachers*

### **1. Evaluation of Professional Practice**

#### A. Rubrics

- ▶ What rubric is being used (e.g., national model like Framework for Teaching, modified version of national model, locally developed)?
- ▶ Did the district check the alignment of the rubric to the Illinois standards or have evidence of alignment?
- ▶ How many levels are there in the rubrics (e.g., 4, Unsatisfactory, Basic, Proficient, Distinguished)?
- ▶ Are different rubrics used for certified non-teaching positions (like counselors, librarians, etc.)? If so, how were these rubrics developed or chosen?

#### B. Evidence Collection

- ▶ How many formal observations are done? How many informal?
- ▶ How long are the observations?
- ▶ When are the formal observations conducted?
- ▶ Are the observations announced, unannounced, or some of each?
- ▶ How do observers record the evidence (e.g., scripting, checklist on tablet or laptop)?
- ▶ Are peer observers used?
- ▶ Are other types of observers other than school administrators and peers used? If so, what other types are used?
- ▶ How many of the observations are done by each type of observer?
- ▶ Are there specific kinds of teachers whom different kinds of observers rate (e.g., new teachers, foreign language teachers)? If so, describe.

- ▶ What evidence besides observations do you collect to assess performance (e.g., artifacts like lesson plans, student work, PD records; student or parent surveys)? Please describe.
- ▶ How is this evidence collected? For example, is some collected during pre- and post-conferences?

#### C. Decision Processes

- ▶ How is the evidence from the observations and other sources used to determine an overall practice rating?
  - Is the process mechanical or judgmental? For example, do evaluators rate each observation on each rubric element, then use a formula to calculate domain scores, then a final score? Or is it left up to evaluators' judgment how to consider the evidence and come up with a final rating?
- ▶ What guidelines or instructions are evaluators given to help them make the final rating?
- ▶ If multiple raters or observers are used, how are their ratings used?

#### D. Differentiation by Teacher Tenure and Performance

- ▶ How does the process differ for new (untentured) teachers?
- ▶ How does the process differ for tenured teachers who are rated “unsatisfactory” or “needs improvement”?

#### E. Evaluator Training

- ▶ Did all raters take and pass the state-provided (Growth Through Learning) training?
- ▶ Was additional training provided by the district? If so, about how many hours?
- ▶ If the district uses a modified version of the Framework, another rubric, or a locally developed rubric, was additional training provided on district-specific features of the evaluation system?
  - Approximately how many hours of additional training did raters receive?
  - Did this training include:
    1. Practice applying rubrics (e.g., rating videos)?
    2. A certification test at the end of the training?

## F. Inter-rater Agreement

- ▶ Has the district collected any evidence of inter-rater agreement?
  - Such as agreement with each other during certification testing, ongoing monitoring of agreement via spot use of multiple observers, regular use of multiple observers/raters?
- ▶ If so, what level of agreement was found? (Ask for a number if available: either % agreement or some other reliability measure—note that it is unlikely that districts outside CPS will have this.)
- ▶ In addition to the initial training, what is the district doing to promote inter-rater agreement (e.g., refresher training, regular opportunities to discuss rating issues with colleagues, regular practice with videos, group/partner observations)?

## 2. Student Growth

### A. Use of Type I and II vs. Type III Assessments

- ▶ For which teachers' is growth measured using a Type I or II assessment and a Type III assessment (grades and subjects)?
- ▶ For which teachers is growth measured using two Type III assessments (grades and subjects)?

### B. Growth Measured on Type I and II Assessments

- ▶ What Type I and II assessments are used (list by grade and subject)?
- ▶ Which teachers are covered by growth measures based on Type I and II assessments?
- ▶ Approximately what percentage of teachers are evaluated using Type I and II assessments?
- ▶ Do teachers/schools have a choice or flexibility in which Type I and II assessments are used? Are teachers in the same grades/subjects evaluated using the same Type I and II assessments?
- ▶ How is the growth measure(s) based on Type I or II assessments calculated (e.g., a value-added model, simple growth, SLO/SGO)?
- ▶ How are student characteristics such as poverty, English proficiency, and special education status taken into account in calculating the growth measure?
- ▶ Is there any evidence for the precision or reliability of this student growth measure? If so, describe.

C. Growth Measured Using Type III Assessments

- ▶ Which teachers are covered only by growth measures based on Type III assessments?
- ▶ Approximately what percentage of teachers are evaluated based only on Type III assessments?
- ▶ What kinds of Type III assessments are used?
- ▶ How is growth based on Type III assessments calculated?
- ▶ Are student characteristics such as poverty, English proficiency, and special education status taken into account in calculating the growth measure? If so, which factors?
- ▶ Is there any evidence for the precision or reliability of this student growth measure? If so, describe.

D. Are SLOs or SGOs Used? If so:

- ▶ Are there guidelines or models provided to teachers and principals to use in developing SLOs?
- ▶ Is there training for teachers and principals in developing SLOs?
- ▶ Is there training for principals in evaluating whether SLOs have been met?
- ▶ Are SLOs focused on growth rather than attainment?
- ▶ Is there a process or guidelines for ensuring the SLOs are aligned with the school improvement plan or district initiatives?
- ▶ How are SLOs reviewed to ensure rigor and comparability across teachers and schools?
- ▶ Are there guidelines in place for scoring SLOs or evaluating whether the SLO has been met?
- ▶ Are group or shared SLOs used? If so, for which teachers (e.g., grades, subjects)?



E. Data Quality

- ▶ How is the district ensuring a teacher's student rosters are accurate and the correct students are included in a teacher's growth measure?
- ▶ Are there guidelines that define which students are counted for each teacher's growth measure?
- ▶ Is there a minimum number of students used for the growth measures?
  - What is that number?
  - How was that number determined?
- ▶ How long does a student need to be enrolled in or attending a teacher's class for the student to be counted in the teacher's growth measure?
- ▶ How are the following exceptions and complications handled?
  - Shared teacher responsibility for students (e.g., team teaching)
  - Changes in student or teacher assignments during the year
  - Student and teacher absences
  - Missing assessment data for some of a teacher's students

F. Combining Type I, II, or III Measures Into a Growth Rating

- ▶ How are the growth measures using Type I, II, and III assessments combined?
- ▶ What is the formula, algorithm, matrix, or table for combining the measures into an overall growth rating?
- ▶ Does this method differ for different groups of teachers (e.g., those with one growth measure based on Type I and II assessments and one based on a Type III assessment, vs. those whose growth measure is based only on Type III assessments)? (If so describe for each group.)

### **3. Combining Growth and Practice Into an Overall Performance Rating**

- A. How Are the Practice and Growth Measures Combined?
  - ▶ What is the formula, algorithm, or matrix for combining the measures into an overall growth rating?
  - ▶ What are the weights given to student growth and professional practice?
- B. Does This Method Differ for Different Groups of Teachers (E.G., Those With One Growth Measure Based on Type I and II Assessments and One Based on a Type III Assessment, Vs. Those Whose Growth Measure Is Only Based on Type III Assessments)?
  - ▶ If so, describe how the weights, formula, algorithm, or matrix for combining the measures differs for each group.

### **4. Communication**

- A. How Has the New Evaluation System Been Communicated to Teachers?
- B. Has the District Used any of the Following Communication Tools:
  - ▶ A handbook or manual describing how the evaluation process is intended to work.
  - ▶ A website or web page with information about how the process works.
  - ▶ Informational meetings or presentations at the schools.
  - ▶ In-person or internet-based training on how the process works.
  - ▶ One-on-one session with evaluator to explain the process.
  - ▶ A channel for asking questions about the system (e.g., hot line, email address, or web page to submit questions).
  - ▶ Any other form of communication.

*Part 2: Principals and Assistant Principals*  
*(Note need to ask about both principals and APs)*

## **1. Evaluation of Professional Practice**

### **A. Rubrics**

- ▶ What rubric is being used (e.g., state model, regional agency model, national model like Marzano, modified version of national model, locally developed)?
- ▶ Did the district check the alignment of the rubric to the Illinois standards or have evidence of alignment? How was this done?
- ▶ How many levels are there in the rubrics (e.g., 4, Unsatisfactory, Basic, Proficient, Distinguished; Unsatisfactory, Needs Improvement, Proficient, Excellent)?

### **B. Evidence Collection**

- ▶ How many formal observations or site visits are done?
- ▶ How long are observations/site visits?
- ▶ Are informal observations used in the evaluation? Is there a procedure for documenting these observations?
- ▶ How do observers record the evidence (e.g., scripting, checklist on tablet or laptop)?
- ▶ Are there types of observers other than principals (for APs) and district administrators (for principals) used? If so:
  - How many of the observations/site visits are done by each type of observer?
  - Are there specific kinds of principals or APs that different kinds of observers rate (e.g., new principals, struggling principals)?
- ▶ What other evidence besides observations/site visits are used to assess performance (e.g., review of artifacts, staff, student or parent surveys)? Please describe.
  - How is this additional evidence collected and documented?

C. Decision Processes

- ▶ How is the evidence from the observations and other sources used to determine an overall practice rating?
  - Is the process mechanical or judgmental? For example, do evaluators rate each observation on each rubric element, then use a formula to calculate domain scores, then a final score? Or is it left up to evaluators' judgment how to consider the evidence and come up with a final rating?
- ▶ If multiple raters or observers are used, how are their ratings used?
- ▶ How is the principal's self-observation used? Is there a defined method of combining this with evaluator ratings, or is this left to the evaluator's judgment?
- ▶ Is there written documentation on how to combine the information from multiple observations and other data sources? Or is this left up to evaluators' judgment?

D. Evaluator Training

- ▶ Did all raters who rate principals or APs take and pass the state-provided (Growth Through Learning) training related to principal evaluation?
- ▶ Did the district provide any additional training? If so, about how many hours?
- ▶ If the district uses a modified version of the state rubric, or another rubric, was additional training provided on the district-specific features of the evaluation system?
- ▶ Did this training include:
  - Practice applying rubrics (e.g., rating videos)?
  - A certification test at the end of the training?

E. Inter-rater Agreement (if applicable; note that some districts might have only one evaluator for principals, but most should have more than one for APs)

- ▶ Has the district collected any evidence of inter-rater agreement?
  - Such as agreement with each other during certification testing, ongoing monitoring of agreement via spot use of multiple observers, regular use of multiple observers/raters?
- ▶ If so, what level of agreement was found? (Ask for a number if available: either % agreement or some other reliability measure—note that it is unlikely that districts outside CPS will have this.)
- ▶ In addition to the initial training, what is the district doing to promote inter-rater agreement (e.g., refresher training, regular opportunities to discuss rating issues with colleagues, regular practice with videos, group/partner observations)?

F. In What Other Ways (Besides any Revealed by Answers to the Questions Above) Does the Process for Evaluating Assistant Principal Leadership Practice Differ From That Used for Principals?

## 2. **Student Growth** (Note definition in state guidance: “a measureable change in a student’s or group of students’ knowledge or skills between two points in time”)

A. Assessments Used for Measuring Growth

- ▶ What Type I and II assessments are used (list by school grade level and by principal vs. AP)?
- ▶ Do schools have choice or flexibility in which Type I and II assessments are used?
- ▶ How is the growth measure(s) based on Type I or II assessments calculated (e.g., a value-added model, simple growth, meeting a growth goal)?
  - If growth goals are used, how does the evaluator score the growth measure or decide if the growth goal has been met?
  - Are there procedures in place for reviewing growth measures or goals for rigor and comparability across principals and APs?

- ▶ How are student characteristics such as poverty, English proficiency, and special education status taken into account in calculating the growth measure?
- ▶ Is there any evidence for the precision or reliability of this student growth measure? If so, describe.

B. Growth Measured Using Type III Assessments

- ▶ Are Type III assessments used for evaluating any principals or APs? If so, for what types of principals or APs?
- ▶ If Type III assessments are used:
  - What kinds of Type III assessments are used?
  - Do individual principals or APs develop their own growth measures, or is this done by the district?
  - Are there guidelines or models provided for use in developing growth measures or goals based on Type III assessments?
  - Is there training for principals or APs in developing growth measures based on Type III assessments?
  - How does the evaluator score the growth measure or decide if the growth goal has been met?
  - Are there guidelines in place for scoring the growth measure or deciding if the growth goal has been met?
  - Are there procedures in place for reviewing growth measures or goals for rigor and comparability across principals and APs?
  - How are student characteristics such as poverty, English proficiency, and special education status taken into account in calculating the growth measure based on Type III assessments?

### C. Other Outcome Measures

- ▶ Are any of the following academic measures outside of growth measured using Type I, II, or III assessments used to evaluate principals or APs?
  - Attainment measures on academic assessments
  - Cohort-to-cohort improvement measures on academic assessments (i.e., NCLB-type)
  - Sub-group performance on academic assessments (e.g., reducing achievement gaps)
  - Pass rates on AP exams (including by subgroup)
  - 21st century skill assessments, WorkKeys assessments
  - Growth or attainment for EL students
  - Any other
- ▶ Are any of the following non-test measures used to evaluate principals or APs?
  - Student attendance
  - Post-secondary matriculation and persistence
  - Graduation rate
  - Percent on track to graduation
  - 9th/10th grade promotion
  - Truancy
  - Expected/unexcused absences
  - Discipline
  - AP completion rates
  - Dual credit earning rates
  - Other non-test measure
  - How are non-test measures scored (e.g., is growth compared to a standard, is attainment compared to a standard, is an attainment or growth goal set)?

D. Data Quality

- ▶ How is the district ensuring a school's student rosters are accurate and the correct students are included in a principal's or AP's growth measure?
- ▶ Are there guidelines that define which students are counted for each growth measure?
- ▶ Is there a minimum number of students used for the growth measures?
  - What is that number?
  - How was that number determined?
- ▶ What is the minimum length of time a student needs to be enrolled at school in order to be counted in the growth measure?
- ▶ How are the following exceptions and complications handled?
  - Midyear changes in school attended
  - Student absences
  - Missing assessment data for some students

E. Generating a Summative Rating for Student Growth

- ▶ Is your district using the Illinois state principal evaluation model's elements and weights for elementary/middle and/or high schools?
- ▶ Is your district using the Illinois state principal evaluation model's student growth performance levels (i.e., Exceeds Goal, Meets Goal, Minimal Growth, No Growth/Negative Growth)? How are the growth measures combined to determine principal and AP summative growth ratings?
- ▶ If your district is not using the state model:
  - What is the formula, algorithm, matrix, or table for combining the measures into an overall growth rating?
- ▶ How does the method of generating a growth rating differ for different groups of principals or APs? (e.g., new principals, principals at different school levels)?



- F. In what other ways (besides any revealed by answers to the questions above) does the method for evaluating APs using growth differ from that used for principals?

### **3. Combining Growth and Practice Into an Overall Performance Rating**

- A. How are the practice and growth measures combined?
- ▶ Is your district using the Summative Rating Matrix from the Illinois state principal evaluation model to combine the rating of student growth and rating of principal practice into an overall summative rating?
  - ▶ If your district is not using the state model:
    - What is the formula, algorithm, or matrix for combining the measures into an overall growth rating?
    - What are the weights given to student growth and professional practice (and other outcome measures, if not included with growth)?
  - ▶ How does the method of generating an overall summative rating differ for different groups of principals or APs (e.g., new principals, principals at different school levels)?

### **4. Self- Assessment, Goal Setting, and Feedback**

- ▶ Does the process provide for the principal or AP to do a self-assessment based on the professional practice rubrics?
- ▶ Does the process provide for a goal setting meeting between the principal and the evaluator to decide on assessments and other outcome measures and establish student growth targets?
- ▶ How does the process provide for feedback to the principal or AP including identifying areas of strength and areas for growth?

## **5. Communication**

- A. How has the new evaluation system been communicated to teachers?
- B. Has the district used any of the following communication tools:
  - ▶ A handbook or manual describing how the evaluation process is intended to work.
  - ▶ A website or web page with information about how the process works.
  - ▶ Informational meetings or presentations at the schools.
  - ▶ In-person or internet-based training on how the process works.
  - ▶ One-on-one session with evaluator to explain the process.
  - ▶ A channel for asking questions about the system (e.g., hot line, email address, or web page to submit questions).
  - ▶ Any other form of communication.

# Document Review Template

*[District Name]*

*State Region:* \_\_\_\_\_

*Urbanicity:* \_\_\_\_\_

## **Description of [District's] Teacher Evaluation System**

### **Evaluation of Professional Practice**

[Short introductory paragraph that summarizes the district's evaluation of professional practice.]

Evaluation rubrics

[Protocol 1A. - Describe in a paragraph.]

Evidence Collection

[Protocol 1B. - Describe in one to several paragraphs.]

Evaluator Training

[Protocol 1E. 1F. - Describe in one to two paragraphs.]

### Final Practice Rating

[Protocol 1C. 1D. - Describe in one to several paragraphs how the final practice rating is determined—note this does not include growth, etc. This should include discussion on how this differs by tenure and performance.]

## **Evaluation of Student Growth**

[Short introductory paragraph that summarizes the district's evaluation of student growth.]

### Assessments

[Protocol 2A. - Describe in one paragraph.]

### Growth Measurement

[Protocol 2B. 2C. 2D - Describe in one to two paragraphs.]

### Data Quality

[Protocol 2E. - Describe in one paragraph.]

### Final Growth Rating

[Protocol 2F. - Describe in one paragraph.]

## **Final Teacher Rating**

[Protocol 3A. 3B. - Describe in one to two paragraphs.]

## **Communication of Evaluation Process & Results**

[Protocol 4A. 4B. - Describe in one paragraph.]

## **Description of [District's] Principal Evaluation System**

### **Evaluation of Professional Practice**

[Short introductory paragraph that summarizes the district's evaluation of professional practice.]

#### Evaluation rubrics

[Protocol 1A. - Describe in a paragraph.]

#### Evidence Collection

[Protocol 1B. - Describe in one to several paragraphs.]

#### Evaluator Training

[Protocol 1D. 1E. - Describe in one to two paragraphs.]

#### Final Practice Rating

[Protocol 1C. 1F. - Describe in one to several paragraphs how the final practice rating is determined—note this does not include growth, etc. This should include discussion on how this differs by tenure and performance.]

### **Evaluation of Student Growth**

[Short introductory paragraph that summarizes the district's evaluation of student growth.]

#### Assessments

[Protocol 2G. - Describe in one paragraph.]

#### Growth Measurement

[Protocol 2H. - Describe in one to two paragraphs.]

## **Other Outcome Measures**

[Protocol 2I. - Describe in one to two paragraphs.]

### Data Quality

[Protocol 2J. - Describe in one paragraph.]

### Final Growth Rating

[Protocol 2K. 2L. - Describe in one paragraph.]

## **Final Principal Rating**

[Protocol 3A. - Describe in one to two paragraphs.]

## **Self-Assessment, Goal Setting, and Feedback**

[Protocol 4. - Describe in one paragraph.]

## **Communication of Evaluation Processes & Results**

[Protocol 5A. 5B. - Describe in one paragraph.]

# Detailed Case Study Methods

	Case Study District 1	Case Study District 2	Case Study District 3	Case Study District 4	Case Study District 5
<b>Region of the State</b>	North	South	Central	North	Central
<b>Urbanicity</b>	Urban	Rural	Urban	Suburban	Urban
<b>First Year of Planning</b>	2011-12	2011-12	2012-13	2013-14	2011-12
<b>First Year of Implementation</b>	2013-14	2013-14	2013-2014	2014-15	2013-14
<b>Composition of Evaluation Committee</b>	Teachers Union Members and District Administrators	Teachers, Principals, Superintendent, Coordinator of Student Development, Guidance Counselor	Teachers, Certified Staff, Administrators, Board Members, Union Leaders	Teachers, Principals, Superintendent, Union Representatives	Teachers, Certified Staff, Administrators, Board Members, Union Representatives
<b>SIG-Funded District</b>	No	Yes	No	No	Yes
<b>Teacher Evaluation Tool</b>	Modified Danielson	Danielson	Modified Danielson	Modified Danielson	Modified Danielson
<b>Principal Evaluation Tool</b>	PEAC Principal Evaluation Model	PEAC Principal Evaluation Model	Modified PEAC Principal Evaluation Model	Governors' State University Principal Performance-Based Evaluation Instrument	PEAC Principal Evaluation Model

	Case Study District 1	Case Study District 2	Case Study District 3	Case Study District 4	Case Study District 5
<b>Training for Observation Instrument</b>	ISBE and district follow-up	ISBE and district follow-up	ISBE and district follow-up	ISBE and district follow-up	ISBE and district follow-up
<b>Implementing Student Growth Model at Time of Interview for Teacher Eval</b>	Yes	Yes	Yes	No	No
<b>Implementing Student Growth Model at Time of Interview for Principal Eval</b>	Yes	Yes	Yes	Yes	Yes
<b>Growth Measures for Teacher Evaluations</b>	Performance Tasks NWEA MAP	End-of-Course Exams	Teacher-selected assessments including NWEA MAP, benchmark assessments, and others	Yet to be determined	NWEA MAP, Type 3 Assessments for subject areas not covered in NWEA MAP
<b>Number of Informal Observations for New Teachers Per School Year</b>	1	4	2	3	3
<b>Number of Informal Observations for Tenured Teachers Within Year Evaluated</b>	2	4	1	3	2



	Case Study District 1	Case Study District 2	Case Study District 3	Case Study District 4	Case Study District 5
<b>Who Observes Teachers?</b>	Principal or Assistant Principal	Principal, Coordinator of Student Development, Guidance Counselor (required to be certified in state-provided Growth through Learning Training)	District Administrators (evaluating principals or contributing to specialized instructional staff), Principals and Associate Principals (teachers)	District Administrators and Principals	District Administrators, Principal or Assistant Principal
<b>Human Resources Linked to Teacher Evaluation Score</b>	Remediation Plan for tenured teachers, dismissal for non-tenured teachers, determines order of dismissal in case of layoffs/RIFs, not related to compensation or promotion	Remediation Plan for tenured teachers, dismissal if no improvement in allotted time frame, dismissal for non-tenured teachers, not related to compensation/promotion	Remediation Plan with clear steps that support the teacher in developing skills needed and aligned with the district sequence of dismissal policies	Remediation Plan	Remediation Plan, determines order of dismissal in case of layoffs/RIFs, not related to compensation or promotion
<b>Remediation</b>					
<b>Hiring</b>					
<b>Firing</b>					
<b>Retention/Layoffs</b>					
<b>Promotion</b>					
<b>Compensation</b>					

## Case Study Methods

### Participants

A total of six districts were chosen for the case study of the PERA Research Study. Five of the six districts are included in this report. An addendum to this report will incorporate data from the sixth district, which wanted to postpone teacher interviews until August 2014.

PERA Evaluation Case Study Analysis Included in This Report	
Total case study districts	5
Total case study schools	10
Total number of interviews	61

### Procedure

To achieve the goal of gathering a variety of perspectives from districts situated across Illinois that have begun implementing PERA-compliant teacher evaluation systems, a purposive sampling technique was used to select districts for the case study. Districts were chosen from a list provided by the Illinois State Board of Education of Race to the Top Phase III grantees and volunteer districts and were further stratified on the following variables: state region, student demographics, number of students within a district, School Improvement Grant-Funded districts versus non-School Improvement Grant-Funded districts, and urbanicity.

Working with contacts in each district, we developed a convenience sample of interviewees that targeted staff members who were either directly involved in the evaluation system design or were comfortable speaking about the implementation process. When choosing teachers to participate in the case study, we asked district contacts to select at least two teachers from each school with a core-concentration in English or Math and one teacher who taught other subjects. We also requested a mixture of tenured and non-tenured teachers.

## Recruitment

District administrators already participating in other aspects of the PERA Research Study were sent letters by the research team describing the case study. The team arranged a telephone meeting with district administrators to discuss details of the case study sampling process, describe the interview protocols, answer any questions, and create a plan for scheduling interviews. Because interview participants were not compensated, our team made every effort to make participation convenient for interviewees by offering flexible schedules and ensuring the interview lasted an hour or less.

## Interview Procedures

The majority of interviews were conducted by phone. Digital recorders were utilized, with interviewees consent, for transcription purposes. The audio recordings were transcribed in order to conduct qualitative analyses, and the audio files were destroyed once interview transcriptions were completed and verified. In advance of interviews, our team requested that interviewees were provided a quiet, private setting to conduct the interview to ensure confidentiality and no interruptions.

## Measures

Interview guides developed were informed by peer-reviewed literature on teacher evaluation systems, ISBE documentation, recent news articles about state-led efforts to implement teacher evaluation systems within the U.S., and best practices in implementing teacher evaluation systems. Interview questions were open-ended with follow-up probes to elicit further conversation and gain clarification on specific aspects of designing and implementing the PERA-compliant teacher and principal evaluation systems. The interview guides were broken into six sections, which followed the main research questions proposed for this case study. Those sections were:

- ▶ Evaluation Goals/Philosophy
- ▶ Evaluation System Design
- ▶ Training and Quality Control
- ▶ Evaluation Measures/Structure
- ▶ Implementation
- ▶ Use/Impact

A total of six separate interview guides (teacher, principal, district administrator, teacher evaluator, principal evaluator, and union representative) were developed based on the role of the participant.

### Analyses

Transcriptions were uploaded to NVIVO software for qualitative analyses. The team developed an iterative process for coding interview transcripts, resulting in three cycles of analyses: 1.) descriptive, 2.) pattern, and 3.) conceptual model building. Descriptive coding was utilized to summarize passages of qualitative data in short phrases. Examples of descriptive codes are *Evaluation Goals*, *Evaluation Design Process*, and *Impact of Evaluation System*. **This type of coding led to a categorized inventory of the interview data's content and set the groundwork for thematic analysis.** Pattern coding was used to organize thematic patterns that describe phenomena in the qualitative data that relate to specific research questions. Examples of pattern codes are *Ways Districts Have Implemented PERA Compliant Evaluation Systems*, *Challenges to Implementation*, and *Ways to Enhance Implementation*. Following pattern coding, researchers developed conceptual models by linking themes and emergent codes generated from the first two cycles of analysis to create higher level understanding about the structure, features, and implementation of the teacher and principal evaluation systems across districts participating in the case study.

In order to ensure inter-rater reliability when coding interviews, researchers chose one coded **interview from each district and coded each other's interviews for both descriptive and thematic codes.** A Code Comparison query was in NVIVO on the chosen interviews. The query analyzed the percentage of agreement between coders and provided a Kappa coefficient score for each code. **Cohen's Kappa coefficient is a statistical measure used for inter-rater reliability.** Kappa coefficient scores are preferred over percentages by researchers when measuring agreement between coders because Kappa coefficient scores take into consideration the amount of agreement that could be expected to occur by chance. The goals for inter-rater reliability were 1.) at least 80% agreement on any given code within an interview and 2.) 70% of all codes for each interview receiving Kappa coefficient scores that **fell between the "good" agreement (.40-.75) and "excellent" agreement (over .75) ranges.** All coders achieved the set goals. For codes that fell in the poor agreement (below .40) range of Kappa coefficient scores, coders discussed the codes in question and reached consensus on the codes.

# Monitoring Protocol

## Professional Practice

### Teacher Observations/Rubrics

1. What rubric/framework does your district use to assess teachers' professional practice?
  - ▶ Does the district use any specialized or adapted rubrics for any teachers or other staff (i.e., counselor or librarians)? If so, for what teachers/staff? Describe the differences.
2. How many formal and informal observations are conducted each evaluation cycle for teachers and other staff?
  - ▶ How long are the observations? (formal vs. informal)?
  - ▶ Are observations announced and/or unannounced? Which are announced and which are unannounced?
  - ▶ Is there a pre-/post-conference?
  - ▶ Do both formal and informal observations count toward a teacher's final practice rating? Explain.
3. Is the observation (or evaluation of practice overall) process different for different types of teachers (e.g., tenured vs. non-tenured teachers)? How so?
4. Who conducts teachers' observations (e.g., administrators, peer evaluators, district personnel, etc./superintendent or other district personnel)?
  - ▶ Are certain raters assigned to certain teachers (e.g., Director of Bilingual Ed only observe ELL teachers)?

5. Are teacher observers formally trained to observe teachers using the observation rubric?
  - ▶ Who conducts the trainings? How often?
  - ▶ Do the trainings include practice applying the observation rubric?
  - ▶ Do observers have to pass a certification test at the end of the training? What does the certification process entail?
  - ▶ How are new raters, particularly those who start mid-year, trained?
6. How does the district provide professional development to those observed (teachers, staff and staff) in order to help them understand the process?
7. How does the district ensure inter-rater reliability or agreement?
  - ▶ How does the district train observers to apply the observation rubric and rate teachers using the rubric in a consistent manner? How often?
  - ▶ What testing has your district done to measure the consistency of ratings across observers? What were the results?
  - ▶ How does the district continually monitor observer ratings to make sure they are reasonably consistent?
8. How do teachers receive feedback following an observations (verbal/written, post-conference, etc.)? When does this occur?

#### Other Measures (teacher)

9. In what other ways besides observations does **your district assess and rate teachers'** professional practice (e.g., peer reviews, surveys, self-assessments)?
10. How does your district assess non-observable aspects of professional practice (e.g., collaborative relationships, leadership activities, professional development, etc.)?
  - ▶ How are these other measures of professional practice assessed?
  - ▶ What evidence does the observer collect and document?
11. How do other measures of professional practice inform the teachers overall practice rating?

### Principal/Assistant Principal Observations/Rubrics

1. What rubric/framework does your district use to assess principal/assistant principal professional practice?
  - ▶ Does the district use any specialized or adapted rubrics for any principals or assistant principals? If so, for what principals? Describe the differences.
2. How many formal and/or informal observations are conducted each evaluation cycle for principals/assistant principals?
  - ▶ How long are the observations (formal vs. informal)?
  - ▶ What types of activities do raters observe to evaluate principal practice?
    - How often?
    - Do raters observe the same activities for all principals/assistant principals? Or does it vary across schools?
  - ▶ Are observations announced and/or unannounced? Which are announced and which are unannounced?
  - ▶ Is there a pre-/post-conference?
  - ▶ Do both formal and informal observations count toward a principal/assistant principal final practice rating?
3. Is the observation (or evaluation of practice overall) process different for different types of principals vs. assistant principals? How so?
4. Who conducts principal/assistant principal observations (district personnel, etc./superintendent or other district personnel)?
5. Are principal/assistant principal observers formally trained to observe principals/assistant principals using the observation rubric?
  - ▶ Who conducts the trainings? How often?
  - ▶ Do the trainings include practice applying the observation rubric?
  - ▶ Do observers have to pass a certification test at the end of the training? What does the certification process entail?
  - ▶ How are new raters, particularly those who start mid-year, trained?

6. How does the district provide professional development to those observed (principals and assistant principals) in order to help them understand the process?
7. How does the district ensure inter-rater reliability or agreement?
  - ▶ How does the district train observers to apply the observation rubric and rate principals/assistant principals using the rubric in a consistent manner? How often?
  - ▶ What testing has your district done to measure the consistency of ratings across observers? What were the results?
  - ▶ How does the district continually monitor observer ratings to make sure they are reasonably consistent?
8. How do principals/assistant principals receive feedback following an observation (verbal/written, post-conference, etc.)? When does this occur?

#### Other Measures (principal/assistant principal)

9. In what other ways besides observations does your district assess and rate principal/assistant principal professional practice (e.g., peer reviews, surveys, self-assessments)?
10. How does your district assess non-observable aspects of professional practice (e.g., collaborative relationships, leadership activities, professional development, etc.)?
  - ▶ How are these other measures of professional practice assessed?
  - ▶ What evidence does the observer collect and document?
11. How do other measures of professional practice inform the principal/assistant principal overall practice rating?



## **Student Growth** (including assessments used to calculate growth)

### Teachers

1. How is student growth calculated in your district as a part of the evaluation process (student growth percentiles, classroom-level value-added, schoolwide value-added, SLOs, etc.)? One model or multiple?
2. Can you walk through the growth model and explain how it works, with examples?
3. How is student growth measured for teachers in non-tested grades or subjects (e.g., art, music, PE)?
4. If your district uses SLOs:
  - ▶ Who determines the SLOs and on what are they based?
  - ▶ How many SLOs are set for teachers?
  - ▶ How are the SLOs reviewed for rigor and comparability across teachers and schools?
  - ▶ How do you score or evaluate whether an SLO has been met?
5. How are student characteristics such as poverty, English proficiency, and special education taken into account when calculating student growth?
6. What Type I assessments are used to determine student growth for teachers?
  - ▶ For which teachers is growth measured using Type I assessment (list by grade and subject)?
7. What Type II assessments are used to determine student growth for teachers?
  - ▶ For which teachers is growth measured using Type II assessment (list by grade and subject)?
8. What Type III assessments are used to determine student growth for teachers?
  - ▶ For which teachers is growth measured using Type III assessment (list by grade and subject)?
9. Do teachers have a choice or flexibility in which Type I, II, and III assessments are used?
  - ▶ Are teachers in the same grades/subjects evaluated using the same assessments?

### Other Outcome Measures (teacher)

10. In addition to growth, what other outcomes measures does your district use to evaluate teachers (e.g., attainment measures)?

### Principals and Assistant Principals

1. How is student growth calculated in your district as a part of the evaluation process (schoolwide value-added, etc.)? One model or multiple?
2. Can you walk through the growth model and explain how it works, with examples?
3. How are student characteristics such as poverty, English proficiency, and special education taken into account when calculating student growth?
4. What Type I assessments are used to determine student growth for principals/assistant principals?
  - ▶ For which principals and assistant principals is growth measured using Type I assessment (list by grade level, and by principal vs. assistant principal)?
5. What Type II assessments are used to determine student growth for principals/ assistant principals?
  - ▶ For which principals and assistant principals is growth measured using Type II assessment (list by grade level, and by principal vs. assistant principal)?
6. What Type III assessments are used to determine student growth for principals/assistant principals?
  - ▶ For which principals and assistant principals is growth measured using Type III assessment (list by grade level, and by principal vs. assistant principal)?
7. Do principals/assistant principals have a choice or flexibility in which Type I, II, and III assessments are used?

### Other Outcome Measures (principal)

8. In addition to growth, what other outcomes measures does your district use to evaluate principals/assistant principals (e.g., attainment measures, attendance/truancy, graduation rate, grade promotion, discipline, AP completion rates, etc.)?

## Data Quality

### Teachers

1. How does your district link student achievement data to teachers?
  - ▶ What is the process for verifying student rosters?
  - ▶ **How is the district ensuring a teacher's student rosters are accurate and the correct students are included in a teacher's growth measure?**
2. How does the district know if data errors occur and how do they fix them? Is there a process to make changes and/or fix errors?
3. Are there guidelines that define which students are counted for each growth measure?
  - ▶ Is there a minimum number of students used for the growth measure?
    - If so, what is it and how was it determined?
4. **How long does a student need to be attending a teacher's class for the student to be counted in the teacher's growth measure?**
  - ▶ Are there students that are excluded from the measure? If so, under what conditions?
5. How are the following situations handled?
  - ▶ Shared teacher responsibility for students
  - ▶ Changes in teacher assignments
  - ▶ Student absences
  - ▶ Missing assessment data for students

### Principals/Assistant Principals

1. How does your district link student achievement data to principals/assistant principals?
  - ▶ What is the process for verifying student rosters?
  - ▶ **How is the district ensuring a school's student rosters are accurate and the correct students are included in a principal's/assistant principal's growth measure?**
2. How does the district know if data errors occur and how do they fix them? Is there a process to make changes and/or fix errors?
3. Are there guidelines that define which students are counted for each growth measure?
  - ▶ Is there a minimum number of students used for the growth measure?
    - If so, what is it and how was it determined?
4. How long does a student need to be enrolled in the school for the student to be counted **in the principal's/assistant principal's** growth measure?
  - ▶ Are there students that are excluded from the measure? If so, under what conditions?
5. How are the following situations handled?
  - ▶ Changes in principal assignments
  - ▶ Student absences
  - ▶ Missing assessment data for students

## **Combining measures to determine the aggregate/final rating or score**

### Teachers

1. How are the practice, growth, and any other measures combined to determine an aggregate teacher evaluation rating or score?
  - ▶ Describe the exact formula, algorithm, or matrix that your district uses to determine the aggregate rating or score.
  - ▶ Note the weights that the district gives to the practice, growth, and other measures to determine the final rating or score. Also, within each component (meaning practice and growth), please describe how various measures are calculated to determine the **component's weight**.
2. Does the district use the same process described above to determine the final rating for all teachers?
  - ▶ If this process varies for different teachers, describe the differences.
3. How do the results of evaluations inform professional development to ensure that the teachers receive targeted PD to improve weaknesses identified during the evaluation process?

### Principals/Assistant Principals

1. How are the practice, growth, and any additional measures combined to determine an aggregate principal evaluation rating or score?
  - ▶ Is your district using the Summative Rating Matrix from the Illinois state principal evaluation model to combine the rating of student growth and rating of principal/assistant principal practice into an overall summative rating?
    - If so, note any tweaks that your district made to this matrix/model, particularly as they may relate to additional measures (beyond the practice and growth scores).

- ▶ **If your district is not using the state’s matrix or model for combining the measure or tweaked the matrix/model:**
  - Describe the exact formula, algorithm, or matrix that your district uses to determine the aggregate rating or score.
  - Note the weights that the district gives to the practice, growth, and other measures to determine the final rating or score. Also, within each component (meaning practice and growth), please describe how various **measures are calculated to determine the component’s weight.**
- 2. Does the district use the same process to determine the final rating for principals compared to assistant or vice principals?
- 3. If this process varies, describe the differences.
- 4. How do the results of evaluations inform PD to ensure that the principals/assistant principals receive targeted PD to improve weaknesses identified during the evaluation process?

## Communication

### Teachers

1. How does the district communicate the teacher evaluation system to stakeholders?
  - ▶ What various mechanisms are used to communicate the systems?
    - Websites?
    - Newsletters?
    - Emails?
    - Meetings?
    - Webinars?
    - Other?
2. Does the district have a strategic communication plan to keep stakeholders informed? If so, describe the plan.

3. How does the district ensure that the communication strategies are effective and consistent in describing the systems to the stakeholders?
  - ▶ Does the district conduct evaluations of the strategies? If so, describe.
  - ▶ Does the district ask stakeholders for their feedback on the strategies? If so, describe.
4. How does the district use evaluation results or stakeholder feedback to improve its communication strategies?

#### Principals/Assistant Principals

1. How does the district communicate the principal/assistant principal evaluation system to stakeholders?
  - ▶ What various mechanisms are used to communicate the systems?
    - Websites?
    - Newsletters?
    - Emails?
    - Meetings?
    - Webinars?
    - Other?
2. Does the district have a strategic communication plan to keep stakeholders informed? If so, describe the plan.
3. How does the district ensure that the communication strategies are effective and consistent in describing the systems to the stakeholders?
  - ▶ Does the district conduct evaluations of the strategies? If so, describe.
  - ▶ Does the district ask stakeholders for their feedback on the strategies? If so, describe.
4. How does the district use evaluation results or stakeholder feedback to improve its communication strategies?