MEMORANDUM

TO: The Honorable JB Pritzker, Governor
The Honorable Tony McCombie, House Minority Leader
The Honorable Don Harmon, Senate President
The Honorable John Curran, Senate Minority Leader
The Honorable Emanuel "Chris" Welch, Speaker of the House

FROM: Dr. Tony Sanders
State Superintendent of Education

DATE: June 28, 2024

SUBJECT: SR0774 - Teacher Evaluation Research Project Report

The Illinois State Board of Education respectfully submits the Teacher Evaluation Research Project Report to the Governor and General Assembly as urged in Senate Resolution 0774 issued by the 102nd Illinois General Assembly.

This report is transmitted on behalf of the State Superintendent of Education. For additional information, please contact Dana Stoerger, Executive Director, Legislative Affairs at (217) 782-6510 or dstoerge@isbe.net.

cc: Secretary of the Senate
Clerk of the House
Legislative Research Unit
State Government Report Center

Illinois State Board of Education Teacher Evaluation Research Project

Final Report

Benjamin West, Eric Larsen, Melia Repko-Erwin, Juan Gabriel Sanchez, Cong Ye, Jessica Giffin, Angela Johnson, Etai Mizrav, Abena Acheampong, Stephanie Levin, Melissa Arellanes, AnhMinh Nguyen, Richa Arora, and Belema Ibama-Johnson

JUNE 2024

American Institutes for Research® | AIR.ORG

Contents

Exhibits

Executive Summary

In January 2010, the Illinois General Assembly passed the Illinois Performance Evaluation Reform Act (PERA; Senate Bill 315; Public Act 96-0861), reforming state policy regarding principal and teacher evaluation. Prior to 2010, the state mandated that Illinois districts evaluate probationary teachers annually and tenured teachers biennially, without specifying particular requirements. PERA changed the landscape by requiring school districts to implement evaluation systems for teachers, principals, and assistant principals. Under these systems, districts must assign educators one of four summative ratings: "Excellent," "Proficient," "Needs Improvement," or "Unsatisfactory." At least 30% of an educator's rating must be based on assessments of student growth, and up to 70% can be based on observations of professional practice and possibly additional performance indicators chosen by the district. New and underperforming teachers rated as "Needs Improvement" or "Unsatisfactory" must be evaluated annually, and tenured teachers rated "Proficient" or "Excellent" must be evaluated at least every 3 years.

The law also established the statewide Performance Evaluation Advisory Council (PEAC)—comprising educators, principals, superintendents, and other stakeholders—to advise the Illinois State Board of Education (ISBE) on the development and implementation of performance systems.

Senate Resolution 774 (Illinois General Assembly, 2022-c) called on the chairs of PEAC to work together to initiate a research study that examines teacher evaluation in Illinois, gathers feedback from stakeholders, and reviews common practices from other states and their points of convergence and divergence with PERA requirements. The goal of the study is to inform recommendations that could address identified challenges while preserving the core benefits of Illinois's current teacher evaluation system.

ISBE commissioned the American Institutes for Research® (AIR®) to conduct this study. PEAC, ISBE, and AIR worked together to develop the following research questions (RQs) to guide the study:

1. RQ1. What are the **components** of districts' teacher and administrator evaluation systems?
2. RQ2. How do districts differ in their **implementation** of these components, and what factors (e.g., district needs, characteristics of the district) do stakeholders identify as potential reasons for these differences?

3. RQ3. What **benefits** have stakeholders experienced implementing components of the teacher and administrator evaluation system with respect to the following?
    - Obtaining useful feedback
    - Supporting improvements to professional practice
    - Informing policy decisions
    - Adapting to the COVID-19 pandemic
4. RQ4. What **challenges** have stakeholders experienced while implementing components of the teacher and administrator evaluation system with respect to the following?
    - Obtaining useful feedback
    - Supporting improvements to professional practice
    - Informing policy decisions
    - Reporting biases or inequities in system components
    - Adapting to the COVID-19 pandemic
5. RQ5. How do stakeholders suggest addressing reported challenges? What **supports**, resources, and changes are recommended?
6. RQ6. What are the distributions of **educator ratings** across districts and schools and over time, and to what extent do educator ratings vary systematically with respect to the characteristics of educators, districts, and schools?

Data Sources

To answer these RQs, AIR collected and analyzed data from five sources:

- **District evaluation plans.** AIR systematically reviewed a representative sample of 51 Illinois district evaluation plans, identified common components of the plans, assessed variation in components across districts, and summarized the extent to which district evaluation plans align with the provisions of the state's PERA legislation (Illinois General Assembly, 2010), administrative code (Illinois General Assembly, 2022-a; 2022-b), and nonregulatory guidance (ISBE, 2015). For detailed information, see the section titled "Review of District Evaluation Plans."
- **Administrative records of teacher evaluations provided by ISBE.** ISBE provided AIR with evaluation ratings ("Excellent," "Proficient," "Needs Improvement," or "Unsatisfactory") for Illinois public school teachers and school-based administrators for the 2013–14 through 2022–23 school years. ISBE also provided AIR with information on the race/ethnicity, gender, district and school codes, years of experience, and position name and code for

these educators. AIR used these data to analyze the distribution of educator ratings across different districts and schools over time, and to see if the ratings differed based on the characteristics of educators, districts, and schools. For detailed information, see the section titled "Educator Evaluation Ratings: Patterns and Trends."

- **Statewide survey of PK–12 teachers and administrators.** AIR developed a 15-minute survey in collaboration with PEAC and ISBE. The survey was administered to a representative, statewide sample of PK–12 teachers (both tenured and nontenured) and administrators (principals and vice principals). By collecting standardized information from a representative sample, the AIR research team was able to draw conclusions about how teachers and administrators statewide feel about the evaluation system. For detailed information, see the section titled "Survey Analysis."
- **Focus groups with PK–12 teachers and administrators.** AIR conducted in-depth focus groups with a smaller sample of PK–12 teachers and administrators who completed the survey. Unlike the survey, which captured standardized responses from many teachers and administrators to make general conclusions about the experiences of all educators in Illinois public schools, the focus groups allowed participants to engage in detailed conversations and provide deeper insights about teacher evaluation in the state. These conversations explored implementation experiences across districts, perceived benefits and challenges, and potential recommendations to address challenges. Although the focus group participants were volunteers whose views may not have been representative of the full population of teachers and administrators in Illinois, their contributions offered detailed insights that helped clarify the survey results. Excerpts in the transcript were coded and tagged to their respective RQs. Additional codes were created to capture unexpected patterns in the data. For detailed information, see the section titled "Interview and Focus Group Analysis."
- **Interviews with former PEAC members and PERA legislators and focus groups with joint committee members.** To provide additional information regarding the purpose and design of PERA at the state and district levels, former PEAC members and legislators involved in writing the original PERA legislation (2010) were invited to participate in interviews, and members of selected districts' PERA joint committees were invited to participate in focus groups. Questions for these interviews and focus groups were developed in collaboration with PEAC and ISBE. These interviews and focus groups complemented data collected from administrators and teachers by providing historical and systems-level perspectives on PERA implementation. Interview and focus group data were systematically coded to identify overall themes and to compare patterns between different groups of participants. For detailed information about these analyses and findings, see the section titled "Interview and Focus Group Analysis."

Methods

After analyzing each of these data sources individually, triangulation methods were employed to develop an integrated understanding of the implementation of educator evaluation systems in Illinois, comparing findings from the survey, interviews, focus groups, evaluation ratings, and evaluation plans. This mixed-methods approach allowed us to identify patterns across data sets, including similarities and differences, thereby enhancing the reliability and robustness of this study's final conclusions. By documenting and supporting key findings with evidence from across multiple data sources, AIR aimed to support a comprehensive and nuanced understanding of the opportunities and challenges faced by Illinois educators in their performance evaluations. For a detailed discussion about these methods, see the section titled "Triangulation."

Findings

In this section, key findings for each research question are summarized. For a detailed discussion of these findings, including supporting evidence, see the section titled "Triangulation."

RQ1. What are the components of districts' teacher and administrator evaluation systems?

The district evaluation plans that were reviewed generally assigned a weight of 70% to professional practice and 30% to student growth in determining an educator's summative evaluation rating. In the survey, educators reported that student growth was assessed using a mix of standardized and locally developed tests. In subsequent focus groups, educators reported that professional practice was commonly measured using the Danielson Framework for Teaching (2007) or Marzano Teacher Evaluation Model (2013a). Most plans included professional development for those rated "Needs Improvement" and "Unsatisfactory," but fewer specified pathways for "Proficient" or "Excellent" educators to access targeted development opportunities based on their performance evaluation.

RQ2. How do districts differ in their implementation of these components, and what factors (e.g., district needs, characteristics of the district) do stakeholders identify as potential reasons for these differences?

Survey data indicated that nearly all educators (97.6%) who participated in the 2022–23 performance evaluation cycle reported being observed at least once as part of that process. Focus groups and interview participants indicated that some districts have customized their observation rubrics to address the needs of teachers in different roles and to allow educators to focus on specific professional practices. Although the educators we surveyed and interviewed shared that they valued pre- and postconferences, they also indicated that access to coaching

opportunities was limited, with tenured teachers typically receiving less coaching than their nontenured counterparts.

RQ3. What benefits have stakeholders experienced implementing components of the teacher and administrator evaluation system with respect to the following: obtaining useful feedback, supporting improvements to professional practice, informing policy decisions, and adapting to the COVID-19 pandemic?

Most survey participants characterized evaluative feedback as actionable for professional growth, with 33.4% describing it as "Extremely Actionable" and 55.9% describing it as "Somewhat Actionable." Several focus group participants emphasized the constructive nature of the suggestions and tools shared by their evaluators. However, different groups of educators reported varying experiences in the survey. For example, a higher proportion of nontenured educators (38.3%) and educators who were not in city schools (35.3% for suburb, 41.0% for town, and 35.7% for rural) described evaluative feedback as "Extremely Actionable" than did tenured educators (28.3%) and educators in city schools (24.0%).

RQ4. What challenges have stakeholders experienced while implementing components of the teacher and administrator evaluation system with respect to the following: obtaining useful feedback, supporting improvements to professional practice, informing policy decisions, reported biases or inequities in system components, and adapting to the COVID-19 pandemic)?

Educators generally felt confident in their understanding of performance evaluation systems, with 52.1% of survey respondents feeling "Extremely Confident" and 39.2% feeling "Somewhat Confident." However, in focus groups, several teachers and administrators described the performance evaluation processes as "burdensome" and "time consuming." Administrators' large evaluation caseloads, an influx of new teachers, the extensiveness of district evaluation processes, and the time required to complete them were cited by several evaluators and educators as reasons for viewing the process as a matter of compliance, rather than an opportunity to enhance teaching and learning.

Survey respondents expressed that the overall evaluation system measured their professional practice generally accurately, with 39.5% describing it as measuring their professional practice "Extremely Accurately" and 50.0% describing it as measuring their professional practice "Somewhat Accurately." However, educators' perceptions of the system's accuracy in measurement differed somewhat by their race/ethnicity, the geographic locale[1] of their school,

[1] The National Center for Education Statistics (NCES) assigns schools to four locales: city, suburban, town, and rural. "City" includes territory inside an urbanized area and inside a principal city. "Suburban" includes territory outside a principal city and inside an urbanized area. "Town" includes territory inside an urban cluster. "Rural" includes census-defined rural territory (Geverdt, 2018).

and the socioeconomic background of students in their schools. For example, a lower percentage of Black or African American educators (24.7%) reported that the system measured their professional practice "Extremely Accurately" compared to 41.4% of White educators. Similarly, the percentage of educators in city schools who reported they were measured by the system "Not at All Accurately" (18.8%) was higher than educators in other locales. Finally, a higher percentage of educators (44.2%) in schools in the 1st (lowest) quartile for enrollment from low-income families reported system measurement as "Extremely Accurately" than educators in schools in the 4th (highest) low-income quartile (32.2%). In focus groups, several teachers, administrators, and joint committee members suggested that differences in evaluator experience and training may lead to inconsistencies in how accurately teachers are evaluated.

When surveyed about the fairness of the formal observation component, 50.8% of educators responded that it measured their professional practice "Extremely Fairly" and 42.0% as "Somewhat Fairly."[2] However, educators' views on the fairness of the observation component differed based on their race/ethnicity and geographic locale. Markedly lower proportions of Black or African American (30.6%) than White (53.1%) educators and lower proportions of educators in city schools (35.3%) than suburban (54.9%), rural (56.1%), and town (58.4%) schools reported that the observation component measured their professional practice "Extremely Fairly." Several interview and focus group participants voiced concerns about subjectivity and scheduling challenges, suggesting these may have contributed to disparities in how fairly educators are evaluated.

When surveyed about the fairness of the student assessment component, 48.8% of educators responded that it measured their professional practice "Somewhat Fairly" and 34.9% as "Extremely Fairly."[3] However, perceptions of the student assessment component's fairness varied, with a lower percentage of educators describing it as reflecting their practice "Extremely Fairly" in city (22.5%) schools than in rural (43.7%) schools; a lower percentage of CPS educators described the student assessment component as measuring their professional practice "Extremely Fairly" (22.9%) than in other districts (39.6%). A lower percentage of educators in schools with the lowest (1st quartile) student-baseline proficiency rates in English Language Arts (ELA) (29.8%) and math (30.9%) described the student assessment component as measuring their professional practice "Extremely Fairly" compared to educators in schools with the highest (4th quartile) ELA (46.1%) and math (42.9%) baseline proficiency rates. During focus group discussions, educators expressed concerns about potential biases in standardized tests

[2] The survey used "formal observation component" as a synonym for "professional practice component." Because survey respondents might not have considered these to be synonyms, we report survey findings using the language respondents encountered.

[3] The survey used "student assessment component" as a synonym for the "student growth component." Because survey respondents might not have considered these to be synonyms, we report survey findings using the language respondents encountered.

and their alignment with educational practices, especially for those teaching students who require specialized support. Additionally, both teachers and administrators who participated in focus groups expressed concerns with student learning objectives (SLOs)[4], describing them as time consuming, susceptible to manipulation, and not significantly contributing to teacher development or student learning. As one special education teacher who participated in a focus group explained, "standardized tests don't show if the student has grown in the areas that their IEPs are . . . when it comes to using that as a measure of whether or not I'm doing my job, that doesn't go together really well."

During focus group discussions, several teachers reported that these challenges with the evaluation contributed to increased job-related stress and diminished trust in evaluators. Participants in the focus group discussions further highlighted challenges related to the COVID-19 pandemic, which disrupted the cadence of the evaluation cycle.

RQ5. How do stakeholders suggest addressing reported challenges? What supports, resources, and changes are recommended?

Although the majority of respondents reported that getting support to improve practice was "Not Difficult at All" (57.9%) or "Somewhat Difficult" (27.7%), a sizable contingent described it as "Extremely Difficult" (14.4%). A closer look at the survey responses shows that views differed among respondents based on school locale and the baseline ELA proficiency rates of the school's students. For example, a substantially higher percentage of educators in city schools (27.8%) than in suburban (11.6%), town (9.0%), and rural (2.9%) schools reported that it was "Extremely Difficult" to get support. Similarly, 22.5% of educators in schools with the lowest (1st quartile) baseline ELA proficiency rates expressed that obtaining support was "Extremely Difficult" compared to 9.2% of educators in schools with the highest (4th quartile) ELA proficiency rates.

To better align the evaluation system with opportunities for professional growth, several focus group participants suggested that districts use instructional coaches to informally observe teachers and provide nonevaluative feedback. In focus group discussions, several teachers and administrators recommended instituting more frequent, nonevaluative observations of professional practice. To address the time and workload barrier to providing support, they also suggested that districts or schools focus their efforts on a few specific domains of the observation rubric. Educators also suggested increased attention to relational elements, such as how teacher-evaluator relationships can support more specific or holistic feedback.

[4] SLOs are measurable targets for student progress over a defined period, typically a school year, in a given content area and grade level.

RQ6. What are the distributions of educator ratings across districts and schools and over time, and to what extent do educator ratings vary systematically with respect to the characteristics of educators, districts, and schools?

In the 2022–23 school year, 51,843 teachers and 2,246 principals/assistant principals received an evaluation rating. Most teachers (97.2%) and principals/assistant principals (96.8%) received "Excellent" or "Proficient" ratings. The distribution of evaluation ratings differed by educator race/ethnicity. For instance, Black or African American teachers received less favorable ratings than their White counterparts, with only 22.1% of Black or African American teachers receiving an "Excellent" rating compared to 50.0% for White teachers. Similar patterns were observed for principals and assistant principals, with Black or African American (27.0%) and Hispanic or Latinx (20.0%) principals and assistant principals receiving ratings of "Excellent" less frequently than their White peers (51.2%).

Black or African American and Hispanic or Latinx teachers consistently received lower ratings than White teachers, regardless of whether they had 1 year, 2–3 years, 4–5 years, or 6 or more years of teaching experience. For example, among first-year teachers, 10.8% of Black or African American teachers and 15.0% of Hispanic or Latinx teachers were rated as "Excellent," compared with 24.0% of White teachers. Conversely, 12.6% of Black or African American teachers and 9.3% of Hispanic or Latinx first-year teachers were rated as "Needs Improvement" or "Unsatisfactory," compared with 4.6% of White first- year teachers.

The distribution of evaluation ratings also differed by school and district characteristics. In schools with the highest (4th quartile) proportions of low-income students, a lower percentage of teachers (26.3%) received an "Excellent" rating compared to teachers (55.4%) in schools serving the lowest (1st quartile) proportions of low-income students. In districts with the highest (4th quartile) proportions of White teachers, a higher percentage of teachers (57.7%) received ratings of "Excellent" than did teachers (38.0%) in districts with the lowest (1st quartile) proportions of White teachers. During interviews and focus groups, an early contributor to PERA and a member of a joint committee raised concerns about the current evaluation system's ability to measure performance reliably and fairly, particularly for Black or African American and Hispanic or Latinx teachers in schools serving low-income communities. These participants suggested that the system does not adequately account for systemic barriers that may adversely impact the evaluations of Black or African American and Hispanic or Latinx educators serving at-risk populations.

Potential Recommendations for Consideration

PEAC is charged with developing recommendations to continuously improve the policy and implementation of educator evaluation using the findings from this study (Illinois General Assembly, 2024-b). Based on these integrated findings, AIR identified the following potential

recommendations for PEAC to consider to inform a set of policy modifications and statewide supports that address any challenges identified while preserving the core benefits of PERA. These potential policy recommendations have been grouped into two categories: potential recommendations related to the implementation of Article 24A of the Illinois School Code (Illinois General Assembly, 2024-a) and potential recommendations related to revisions to Article 24A of the Illinois School code itself.

Potential Recommendations Related to the Implementation of Article 24A of the Illinois School Code:

- Minimize the burden of the evaluation system while prioritizing actionable feedback and professional growth for all educators.
- Calibrate evaluators and multiple observers.
- Support diverse and culturally inclusive evaluation practices.
- Foster positive relationships and trust between administrators and teachers.

Potential Policy Recommendations Related to Revisions to Article 24A of the Illinois School Code:

- Reconsider the use or incorporation of student growth.

Potential Recommendations Related to the Implementation of Article 24A of the Illinois School Code:

Minimize the Burden of the Evaluation System While Prioritizing Actionable Feedback and Professional Growth for All Educators.

Educators in the focus groups shared concerns that educator evaluation is often time consuming, and the quality or amount of feedback or support may not be worth the time investment for every aspect of the evaluation system. PEAC may consider recommendations to prioritize components of the evaluation system that offer greater opportunities for actionable feedback, such as classroom observations, over more time-intensive processes, such as SLOs, which may not provide educators with actionable feedback. In developing recommendations, PEAC may want to consider supports that will help Illinois to prioritize actionable feedback and professional growth, so that evaluators and educators are more likely to view the system as an opportunity to enhance teaching and learning, rather than viewing the process as a matter of compliance.

Calibrate Evaluators and Support Multiple Observers.

Currently, evaluators must undergo training and retraining to become evaluators. The analysis of evaluation ratings data identified systematic differences in educator ratings by the race and ethnicity of teachers, which suggests that evaluators may need additional calibration. One potential consideration is to revisit the cadence and content of evaluator training to reduce subjectivity and bias and improve interrater reliability among evaluators. Existing evaluator training could be strengthened by including video examples highlighting potential areas of bias or subjectivity. To help provide more frequent feedback opportunities, PEAC may consider recommending guidance and support for districts to offer multiple observers, including teacher leaders. This could provide multiple perspectives on a teacher's practice and give some flexibility to allow educators with the same content or subject area as the teacher being observed to observe and give feedback. Training additional observers might also help ease the time burden for administrators, and it could provide teachers with feedback specific to their content or subject area.

Support Diverse and Culturally Inclusive Evaluation Practices.

In the evaluation ratings data, educator evaluation ratings varied across racial and ethnic lines, and in the survey data, concerns regarding the fairness and accuracy of the evaluation system differed substantially by race/ethnicity and locale. This may suggest that the implementation of the evaluation system could be continuously improved by embedding more diverse and culturally inclusive evaluation practices in the system. Embedding principals of diversity, equity, and inclusion in training for evaluators or other professional learning opportunities could support diverse and culturally inclusive evaluation practices. PEAC may also consider recommending that ISBE help address concerns of bias in the evaluation system by regularly reporting on teacher evaluation ratings by teacher and school characteristics in a way that also protects teacher privacy. Producing annual tables similar to those presented in the Educator Evaluation Ratings Patterns and Trends section of this report may help alleviate perceptions of bias or spur action to address any identified bias from districts, the general assembly, or others in the education community.

Foster Positive Relationships and Trust Between Administrators and Teachers.

Establishing trust is essential for any coaching relationship between teachers and school leaders (Bryk & Schneider, 2002; Cranston, 2011; Tschannen-Moran, 2001). In focus groups, some teachers expressed increased strain on the relationship between teachers and evaluators. Several teachers and administrators highlighted relationships as a key part of successful evaluation systems, which entailed informal coaching and walk-throughs. One consideration for PEAC is a statewide and/or regional professional learning community focused on strategies to

build trust and positive relationships between teachers and evaluators with opportunities to share examples for different building sizes.

Potential Recommendations Related to Revisions to Article 24A of the Illinois School Code:

Reconsider the Use or Incorporation of Student Growth.

This report's review of district evaluation plans and focus groups with teachers and school leaders suggest that most districts in Illinois use SLOs to measure student growth. Reconsidering the state's approach to student growth is a resonant theme culled from data. Some examples of how the state might change its use of student growth measures to evaluate teachers include the following:

- Eliminate the student growth requirement from evaluations.
- Reduce the percentage weight of student growth in the overall evaluation score.
- Offer a statewide student-growth model, such as a VAM or SGP, for teachers of Grades 4–8 ELA and math, and eliminate the student growth requirement for all other teachers.
- Embed the instructional practices from the SLO process—such as analyzing student data, setting growth targets, and using data to inform instruction—into the professional practice rubric. This would eliminate the separate weighted measure of student growth but maintain the emphasis on the data analytic practices that inform instruction. This approach could also include a focus on the impact on student learning within the professional practice framework.

The state of Illinois commissioned this evaluation of PERA implementation to inform potential policy recommendations and improve the overall evaluation system for educators across the state. This study provided a broad set of evidence, including a policy scan of teacher evaluation systems across the United States, a survey of educators who have experienced PERA implementation, interviews and focus groups with early contributors to legislation as well as educators involved in its implementation, a review of a wide cross section of district evaluation plans, and an analysis of educator ratings across the state. The potential policy recommendations presented here are based on the findings from a synthesis of these data and are intended to inform PEAC's recommendations regarding further changes to the overall teacher evaluation system in the state of Illinois.

Introduction and Teacher Evaluation Policy Scan

In January 2010, the Illinois General Assembly passed the Illinois Performance Evaluation Reform Act (PERA; Senate Bill 315; Public Act 96-0861), reforming state policy regarding principal and teacher evaluation. Before 2010, the state mandated that Illinois districts evaluate probationary teachers annually and tenured teachers biennially without specific requirements. PERA changed state policy by requiring school districts to implement evaluation systems with specific requirements for teachers, principals, and assistant principals. Under these systems, districts must assign educators one of four summative ratings: "Excellent," "Proficient," "Needs Improvement," or "Unsatisfactory," and at least 30% of a teacher's summative rating must be based on assessments of student growth. New and underperforming teachers rated as "Needs Improvement" or "Unsatisfactory" must be evaluated annually, and tenured teachers rated "Proficient" or "Excellent" must be evaluated at least every 3 years.

The law also established the statewide Performance Evaluation Advisory Council (PEAC), consisting of educators, principals, superintendents, and other stakeholders, to advise the Illinois State Board of Education (ISBE) on developing and implementing performance evaluation systems. Twelve years after PERA's enactment, in 2022, the Illinois General Assembly passed Senate Resolution 774, recommending that PEAC conduct a research study on the implementation of teacher evaluation in Illinois, including a review of common state practices and their points of convergence and divergence with PERA requirements.

In response to the directive, this chapter provides an overview of state teacher evaluation policies by describing measures that are commonly implemented and reviewing evidence about the extent to which these measures may contribute to valid, reliable, and transparent evaluations of teacher performance. For each measure, points of convergence and divergence between PERA requirements and commonly implemented evaluation approaches in other states are identified; how states vary in their requirements related to district flexibility in developing and/or determining the criteria for teacher evaluation systems is examined; and the use of teacher evaluation measures for supporting teacher development and, ultimately, student improvement is considered.

The analysis yielded four key findings:

- Common among many states, including Illinois, is the use of a combination of evaluation measures. In Illinois, this includes classroom observations and student growth measures such as student learning objectives (SLOs). Districts in Illinois place substantial emphasis on classroom observations, which is common across many states. This section expands on and provides examples of different approaches adopted by states.

- There are various approaches to statewide teacher evaluation policies:
  - Ten states mandate a comprehensive model to which all districts must adhere (e.g., Ohio, Georgia, Washington).
  - The District of Columbia and 26 states, including Illinois, allow districts flexibility within some state parameters (e.g., Indiana, Iowa, Maryland).
  - Fourteen states set minimum state criteria and allow districts flexibility while also offering an "opt-in" default option (e.g., Massachusetts, Tennessee, Texas).
- Teacher evaluations can be useful in providing district and school leaders with valuable information regarding teachers' professional development needs and in providing teachers with feedback to inform their practice (Hunter, 2022; Lovison & Taylor, 2018). PERA illustrates this best practice well, with targeted feedback and detailed improvement plans being a part of the evaluation process for teachers who need improvement. The approach emphasizes multiple observations for new and low-performing teachers and mandates rapid communication about the evaluations.

In the remainder of this chapter, the components of PERA and related expectations for Illinois's school districts are described in greater detail. Next, a set of commonly implemented evaluation approaches: measures used for evaluation, state policies on mandating state or local evaluation systems, and states' approaches to providing feedback, mentoring, and professional development to educators are examined. To place the approach used in Illinois in context, three states with similar characteristics in terms of region, demographics, and student outcomes—Michigan, Ohio, and Wisconsin—are considered.

The Illinois Performance Evaluation Reform Act

The Illinois Performance Evaluation Reform Act, or PERA, requires districts in Illinois to design and implement performance evaluation systems that assess teachers' and administrators' professional practices and contributions to student growth. Based on multiple measures of educator performance and student outcomes, PERA requires all teachers and administrators to be rated in one of four categories (i.e., "Excellent," "Proficient," "Needs Improvement," or "Unsatisfactory") regardless of tenure status (Illinois General Assembly, 2010). This rating is used not only to assess teachers, but also to determine the frequency of evaluation. The Illinois Administrative Code further specifies that formal evaluations of new and underperforming teachers (rated "Needs Improvement" or "Unsatisfactory" in any one year) must occur annually, and formal evaluations of tenured teachers must occur at least every 3 years if they are rated "Proficient" or "Excellent" (Illinois General Assembly, 2022-a).

The measures included in most Illinois district performance evaluation systems include professional practice and student growth.

- **Professional practice.** Tenured teachers with an "Excellent" or "Proficient" rating from their last performance evaluation must be observed at least twice during the 3-year evaluation cycle, with at least one formal observation of professional practice. If a tenured teacher received a "Needs Improvement" or "Unsatisfactory" rating, they must be observed at least three times in the school year following that rating, with at least two of those observations being formal. Nontenured teachers should be observed at least three times a year, with at least two formal observations (Illinois State Board of Education, 2022). In addition to requiring that teacher observations be based on standards of effective practice, a notable aspect of PERA is that it calls for evaluators to be trained and prequalified to conduct observations, collect evidence, and provide helpful, timely feedback.
- **Student growth.** The biggest change for districts upon the enactment of PERA was the requirement that student growth be a factor in every evaluation. The law (Illinois General Assembly, 2010) and associated regulations (Illinois General Assembly, 2022-a; 2022-b) require that student growth represent a "significant factor" in evaluations, amounting to at least 30% of the overall evaluation.

When established, PERA required each district to establish a joint committee to develop an educator evaluation plan that complies with PERA requirements. ISBE's non-regulatory guidance states that "it is the intent of PERA that those individuals closer to the evaluation process (i.e., local administrators and teachers) should be the members of the Joint Committee" (ISBE, 2015 p. 28). Members of each joint committee were determined by the district and its teachers (or representatives of teachers' bargaining units), with equal numbers of committee members selected by both the district and teachers or their representatives. School districts were allowed to develop their own teacher evaluation systems that met minimum standards mandated by state rules. However, if a district joint committee failed to agree on a plan for the district within 180 days of its first meeting, the district was required to implement the state's model evaluation plan. The Illinois Administrative Code (see subpart C: 50.200-50.230 of Illinois General Assembly, 2022-a) outlines the state's model evaluation plan, in which 50% of a teacher's evaluation rating is based on student growth, and student growth is measured using SLOs.

As noted above, PERA also established PEAC, which encompasses a diverse range of educational leaders from Illinois, including teachers, principals, superintendents, and directors from various school districts and educational and policy organizations. Senate Resolution 774 (Illinois General Assembly, 2022-c) subsequently called for "PEAC to initiate and complete a research study . . . to examine the implementation of teacher evaluation in Illinois, gather

feedback from stakeholders state-wide, and review best practice from other states, [and] to use that research to inform a set of policy recommendations that would address any identified challenges while preserving the core benefits of PERA's robust evaluation system."

Approximately 5 years after PERA was passed, Congress's reauthorization in 2015 of the Elementary and Secondary Education Act of 1965 as the Every Student Succeeds Act (ESSA) eliminated the federal requirement for statewide teacher evaluation systems (Ross & Walsh, 2019, p. 3). ESSA also eliminated federal requirements for teacher quality and effectiveness, including requirements for teachers to demonstrate subject matter competency on a test or through coursework completion in middle/high schools, and eliminated prohibitions against the staffing of teachers holding "emergency" or "provisional" certification in Title I schools. Nevertheless, ESSA required states to determine whether low-income and minority students are disproportionately taught by ineffective, inexperienced, or out-of-field teachers (Barone, 2017). States had to develop and implement plans to address these disparities if they were found, effectively compelling states to maintain or establish criteria for defining "ineffective" teaching, with most states opting to continue using their pre-ESSA criteria (Mizrav, 2019).

Despite the relaxation of federal requirements in 2015, approximately one quarter of all states, including Illinois, either continued to develop or did not scale back their evaluation systems (Close et al., 2020). Notably, the Illinois General Assembly did not scale back the requirements that it had established for teacher evaluation under PERA in 2010.

Review of State Teacher Evaluation Policies

This section provides a review of state teacher evaluation policies, with special attention to the types of measures that states use to assess teachers' professional practice and contributions to student growth. It begins with a description of the properties considered in determining the usefulness and value of various measures of teacher evaluation then presents common components of state teacher evaluation systems, reviews states' requirements regarding district flexibility in these evaluation systems, and considers states' use of evaluations to support teachers' professional growth.

Transparency, Validity, and Reliability in Teacher Evaluation Measures

Educational measurement research focuses on properties such as transparency, validity, and reliability to help determine the suitability of a measure for a specific use, such as teacher evaluation (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014).

Transparency concerns the clarity and openness with which information about the purpose, procedures, and outcomes of a measure are shared with relevant stakeholders. If educators

have a clear understanding of how a student growth measure is calculated, or how teachers' classroom practices are scored on an observation rubric, those measurement tools may be characterized as transparent. Transparency helps teachers understand the practical application of the evaluation results for their professional growth and development (Marzano, 2013b).

Validity refers to the extent to which evidence and theory support a specific interpretation or use of a measure. For instance, if the measure is designed to provide a valid assessment of a teacher's effectiveness at improving student learning, it is reasonable to expect the measure to show a correlation with student achievement gains. If the measure has little or no correlation with student achievement gains, but a high correlation with student characteristics (such as the student's achievement in the prior year) or teacher characteristics (such as teacher race/ethnicity or gender), the measure will have low validity and would be biased. A valid measure would also properly assess educator quality in different contexts (American Educational Research Association, 2014).

The **reliability** of a measure refers to its precision and consistency across repeated applications of the measurement procedure. If different evaluators are using the same observation rubric to rate the same educator, they should arrive at similar ratings for that educator, which demonstrates interrater reliability. Reliability does not guarantee validity, and vice versa: a measure that provides a consistent rating for an educator may not correctly evaluate their performance, and a valid performance evaluation measure may not be reliable. For example, a bathroom scale might consistently display the same weight every time it is used (reliable), but if it is calibrated incorrectly and adds 5 to 10 pounds, a measure of actual weight is not provided (not valid). Conversely, a bathroom scale might show the correct weight when calibrated correctly (valid), but if results vary depending on being placed on different types of flooring, such as plush carpet, soft linoleum, firm but uneven wood, or hard and flat tile, it may show a different weight every time it is used (not reliable).

Developing valid measures, implementing them reliably, and reporting their results transparently presents a significant challenge, particularly for teacher performance evaluation systems. For example, although some student growth models can provide valid measures of teacher contributions to learning (Bacher-Hicks et al., 2014; Chetty et al., 2014; Glazerman et al., 2013; Kane & Staiger, 2008), growth models are often not transparent and are difficult to interpret (Pivovarova et al., 2016). Similarly, growth measures are often unreliable, which is due in part to idiosyncratic influences on student learning unrelated to the teacher (Baker et al., 2010). For instance, the performance of students on standardized tests can be influenced by factors beyond a teacher's control such as family issues, health, or access to resources. Because no single educator evaluation measure is perfectly transparent, valid, and reliable, it is recommended that teacher evaluation systems include multiple measures demonstrating

different strengths to address the limitations inherent in any single evaluation measure (Martínez et al., 2016).

Common Components of Teacher Evaluation Systems

In this section, various measures used to evaluate teacher effectiveness, focusing particularly on the convergence and divergence between Illinois's PERA and practices in other states, are discussed. For each evaluation measure, research findings on its relative transparency, validity, and reliability for use in educator performance evaluation systems are summarized. To these ends, measures that are incorporated in PERA, which include measures of professional practice and student growth, are introduced, followed by a discussion of measures not included in PERA but commonly used in other states.

Measures Converging With PERA

Measure of Professional Practice: Classroom Observations

PERA (Illinois General Assembly, 2010) and the Illinois Administrative Code (Illinois General Assembly, 2022-a) state that tenured teachers with an "Excellent" or "Proficient" rating from their last performance evaluation must be observed at least twice during the 3-year evaluation cycle, with at least one formal observation of professional practice. If a tenured teacher receives a "Needs Improvement" or "Unsatisfactory" rating, they must be observed at least three times in the school year following that rating, with at least two of those observations being formal. Nontenured teachers should be observed at least three times, with at least two formal observations.

The Illinois Administrative Code (Illinois General Assembly, 2022-b) requires school districts to conduct teacher observations using a rubric that adheres to a research-based instructional framework (i.e., addresses planning, instructional delivery, and classroom management) and aligns with Illinois Professional Teaching Standards. The instructional framework used must also align with each teacher's unique roles and responsibilities. Evaluators must participate in prequalification training and then participate in retraining every 5 years to continue evaluating staff. Districts have the option of developing their own prequalification training program.

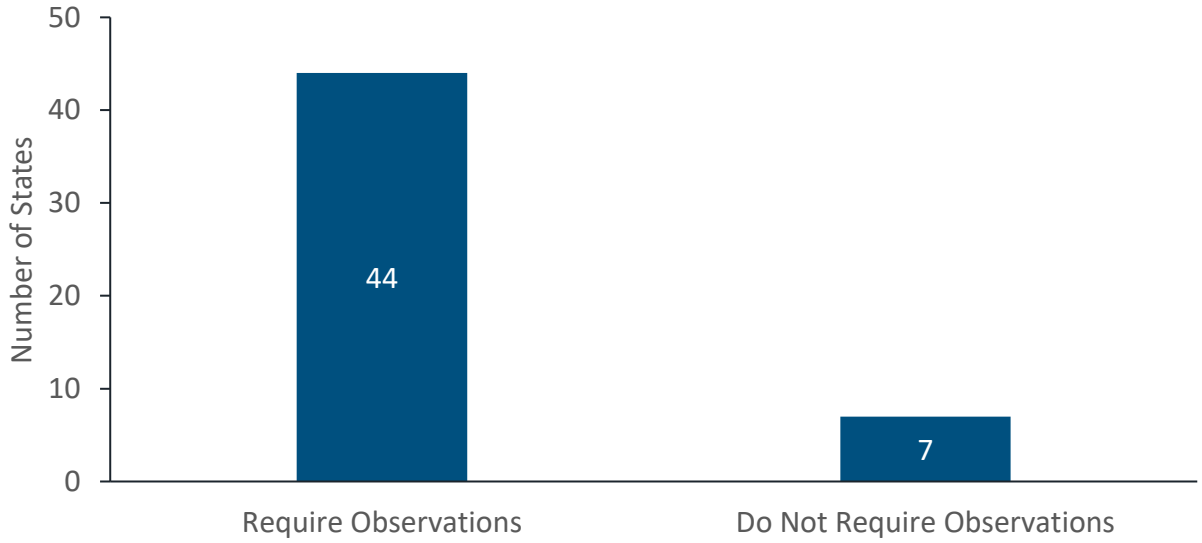
Classroom observations provide a stable and transparent measure that allows for timely formative feedback that teachers could use to improve their practice (Whitehurst et al., 2014). Typically, these observations are conducted by the school principal or assistant principal, in-school teacher leaders, or trained observers from outside the school district. The frequency of observations that a teacher receives can vary significantly, ranging from a single observation to multiple observations throughout the year. Recently, video observations have become more

common, especially during the COVID-19 pandemic (National Council on Teacher Quality [NCTQ], 2022).

Research suggests that well-constructed and well-implemented classroom observation rubrics are predictive of teacher effectiveness, regardless of classroom composition (Bacher-Hicks et al., 2019). Further, professional practice measures, which rely in part on evidence-based rubrics of effective instructional practice, can help teachers build their understanding of good instruction and facilitate improvement by supporting useful, targeted feedback and collaboration (White & Maher, 2024).

Use of classroom observations as a component of Illinois's teacher evaluation system is also consistent with practices in other states. Like Illinois, most states (44) mandate the incorporation of observations into their teacher evaluation systems, with only seven exceptions: Missouri, Montana, Nebraska, New Hampshire, North Dakota, Vermont, and Washington, DC. (See Exhibit 1.)

Exhibit 1. Number of States That Require Classroom Observations for Teacher Evaluations, 2022



Note: Data from National Council on Teacher Quality (2022).

Many commonly used observation measures are based on detailed rubrics to evaluate various aspects of professional practice, making them highly transparent. States and districts often use commercially available rubrics such as the Danielson Framework for Teaching (2007) and the Marzano Teacher Evaluation Model (2013a). Such rubrics support well-structured, easily comprehensible evaluations and offer teachers actionable areas for improvement (Tuma et al.,

2018). For example, the Marzano model simplifies instruction into manageable elements, underscoring practices such as "use preview questions before instruction or a teacher-directed activity," or "explain linkages between mathematical patterns and structure from previous grades/lessons and current content" (Marzano, 2017, p.11). Teachers can clearly understand the exact elements being evaluated and considered and can easily apply these elements to areas for improvement.

Although rubric-based observations are designed to provide transparent, valid, and reliable measures of teacher effectiveness, concerns remain about—and research has provided evidence of—bias in classroom observation measures. For example, using data from the Measures of Effective Teaching study, which was carried out over 2 school years (2009–10 and 2010–11) and across six districts, Steinberg and Garrett (2016) found that teacher observation scores based on the Danielson Framework for Teaching were associated with the prior-year subject-matter achievement of their students. Similarly, using observation ratings of teachers in Chicago Public Schools (CPS) from the 2013–14 and 2014–15 school years, Steinberg and Sartain (2021) present evidence that teacher observation ratings are determined in part by the characteristics of teachers' students, including students' poverty status, prior achievement, and prior-year misconduct. This evidence highlights how teacher observation scores not only reflect teacher effectiveness but are also influenced by student characteristics, underscoring the need for careful consideration of contextual factors that may diminish reliability.

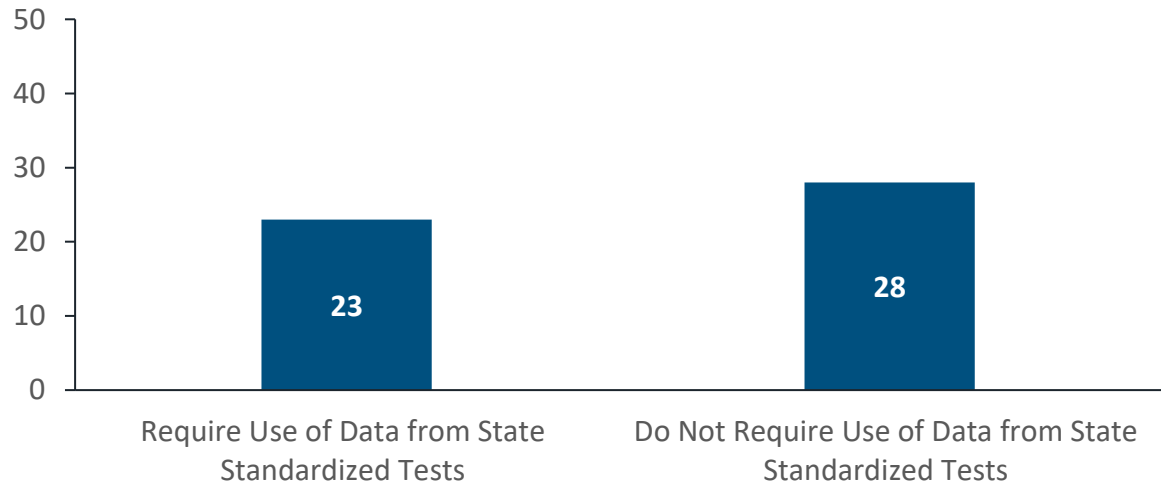
Observation tools often result in most teachers being rated as effective or highly effective (Weisberg et al., 2009). Compared to value-added measures, which are described later in this chapter, classroom observation ratings are clustered more closely together rather than spread out across the possible range of scores (Jacob & Lefgren, 2008). This clustering may imply issues with the reliability of the measure, suggesting that observation tools may fail to differentiate effectively between varying levels of performance. A study examining teacher performance ratings in 24 states showed that, despite some variance in proficiency or below proficiency ratings, less than 1% of teachers in most states are rated "Unsatisfactory" (Kraft & Gilmour, 2017). The survey included in the study also revealed that evaluators believe the number of teachers in their schools performing below "Proficient" is more than triple the number they officially rate as such.

Measures of Student Growth

Illinois requires that a minimum of 30% of a teacher's evaluation be based on student growth. Illinois's approach of having a component of their teacher evaluation system based on student growth on assessments is consistent with national practice. As shown in Exhibit 2, 23 states, including Illinois, explicitly allow the use of data from state standardized tests in teacher evaluations. According to the National Council on Teacher Quality, there has been a gradual

decline in the use of test-based student growth measures for teacher evaluation during the last decade, with 37 states requiring them in 2015, dropping to 28 states in 2019 before the pandemic, and further declining to 23 states after the pandemic (NCTQ, 2022). The decline in the use of test-based student growth measures for teacher evaluation may be due, in part, to Congress's reauthorization in 2015 of the Elementary and Secondary Education Act of 1965 as the Every Student Succeeds Act (ESSA), which eliminated the federal requirement for statewide teacher evaluation systems.

Exhibit 2. Number of States That Explicitly Allow the Use of Data From State Standardized Tests in Teacher Evaluations, 2022



Note: Data from National Council on Teacher Quality (2022).

Assessments Used to Measure Student Growth in Illinois

According to the Illinois Administrative Code (Illinois General Assembly, 2022-a), student growth may be measured using three types of assessments. The first type, commonly referred to as Type I assessments, are standardized assessments that are administered under the same conditions to all students, use a fixed set of rules for scoring responses, and can be administered in different locations and at different times while still producing comparable results. These assessments are scored externally by a nondistrict entity and are applicable statewide or beyond Illinois. Examples of this type are the Illinois Assessment of Readiness, the SAT, and the Northwest Evaluation Association's (NWEA) Measures of Academic Progress (MAP) tests. The second type, known as Type II assessments, are assessments that are developed, adopted, and approved for use by the school district. They are utilized districtwide by all teachers in a specific grade or subject area. The third type, referred to as Type III assessments, are rigorous assessments that align with the curriculum and are determined by the teacher and evaluator to measure student learning in that course. They can include

teacher-created assessments, assessments designed by textbook publishers, student work samples or portfolios, and other locally developed assessments.

Although PERA requires the use of these types of assessments to measure student growth, it allows each district to decide how student growth is measured (Illinois State Board of Education, 2015). In what follows, four types of student growth measures—gain scores, student growth percentiles (SGPs), value-added model (VAM) scores, and SLOs—are described. Before describing these four types of student growth measures, note that the validity and reliability of a student growth measure depends, in part, on the validity and reliability of the student assessment used in the growth measure. If the assessment on which the student growth measure is based is not a valid, reliable measure of the desired student outcomes, then the growth measure itself is unlikely to be a valid, reliable measure of teachers' contributions to student learning.

Gain Scores. Gain scores are a measure of student growth derived by measuring the difference between student scores from two different time points. While this simple measure is both transparent and reliable, there are some issues with this approach. Growth can depend on the initial status of the student rather than the quality of teaching. While the simpler measure may be more transparent than more complex statistical measures, there is limited reliability and validity evidence to support its use for high-stakes decision making (Castellano & Ho, 2013).

Student Growth Percentiles. SGPs use quantile regression to measure how much a student has learned compared with a group of peers in the current school year who had similar test scores in prior years (Doss, 2019; Walsh & Isenberg, 2015). Each teacher's evaluation rating is typically based on either the median or the mean of the SGPs. States such as Colorado and Georgia, as well as Washington, DC, use SGPs (Michigan Department of Education, 2023; Walsh & Isenberg, 2015).

Value-Added Model Scores. VAMs, used in developing value-added scores, improve upon gain-score models and SGPs by accounting for factors (other than the teacher) that influence the rate at which students learn. VAMs typically adjust for differences in student characteristics even beyond prior test scores, considering factors such as a student's poverty, special education, and English language learner status. Guarino et al. (2015) suggest VAMs may do a better job than SGPs of accounting for differences in student characteristics and assessing actual teacher quality. The application of value-added approaches in educational research and teacher evaluation may be understood, for instance, as the difference between students' expected test scores (based on comparisons to students with similar prior test scores and other characteristics) and actual test scores. If all of a teacher's students' actual test scores exceed their expected test scores, the excess growth is attributed to the teacher, and the teacher

would receive an above-average score. Likewise, if all the students in a class do worse than expected, the teacher would receive a below-average score (Opper, 2019).

By accounting for factors other than the teacher that influence the rate at which students learn, VAMs can isolate a teacher's contribution to their students' test score growth (Chetty et al., 2014; Kane & Staiger, 2008). Koedel et al. (2015) reviewed the evidence on VAM and suggested that it can distinguish teachers based on quality meaningfully and reliably in Grades 4–8 English Language Arts (ELA) and math. Koedel et al. noted, "To date, the studies that have used the strongest research designs provide compelling evidence that estimates of teacher value-added from standard models are not meaningfully biased by student-teacher sorting along observed or unobserved dimensions. It is notable that there is not any direct counterevidence indicating that value-added estimates are substantially biased" (Koedel et al., 2015, p. 185).

Although VAMs in Grades 4–8 ELA and math may produce valid measures of teacher effectiveness, the reliability of VAM is relatively low. For example, studies have shown that a teacher rated "Effective" one year may be deemed "Ineffective" the following year (Close et al., 2020; Martínez et al., 2016; Yeh, 2013). The lack of transparency is another significant shortcoming of VAM, which involves intricate analysis typically executed by statisticians. The complexity of VAM may be challenging for teachers to comprehend, and the complexity also makes it difficult for teachers to use VAM effectively for formative improvement and growth (Close et al., 2020; Goldring et al., 2015).

The American Educational Research Association issued a cautionary statement against the use of VAM for high-stakes decisions (American Educational Research Association, 2015). However, Goldhaber and Hansen (2013) note that other available measures pose similar reliability and validity issues. The American Statistical Association (ASA) "endorses wise use of data, statistical models, and designed experiments for improving the quality of education" (2014; p. 1). However, the ASA (2014; p. 1) also notes that "high-level statistical expertise is needed to develop VAMs and interpret their results," and that "estimates from VAMs should always be accompanied by measures of precision and a discussion of the assumptions and possible limitations of the model," particularly if VAMs are used for high-stakes purposes.

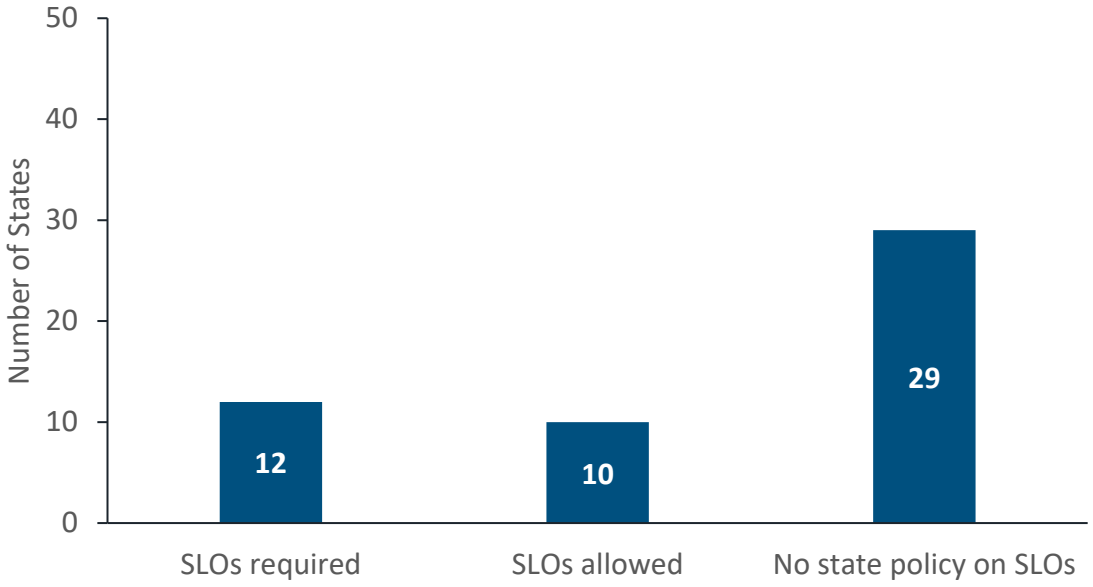
Student Learning Objectives. SLOs are content- and grade/course-specific learning objectives that can measure student learning over a specific time frame, often the school year. Teachers are evaluated based on the extent to which students met the identified objectives. SLOs are frequently used in nontested grades and subjects (Reform Support Network, 2014).

Evidence for the reliability and validity of SLO measures is limited since they take various forms and are not as well established as other measures. However, while evidence is limited, it

suggests that their reliability and validity can be strong when they are implemented rigorously and use concrete data benchmarks (Lin et al., 2020).

As shown in Exhibit 3, 12 states explicitly require use of SLOs in their teacher evaluation systems and 10 states explicitly allow these measures in their policy. The remaining 29 states neither mention nor prohibit the use of SLOs. In nine states, SLOs serve as a substitute for student growth measures for subjects that are not tested (Close et al., 2020; NCTQ, 2022).

Exhibit 3. Number of States That Require or Allow the Use of SLOs in Teacher Evaluations, 2022



Note: Data from National Council on Teacher Quality (2022).

Other Measures of Teacher Effectiveness

Whereas PERA mainly emphasizes incorporating professional practice and student growth measures into teacher evaluations, this section discusses additional evaluation measures used in some states beyond professional practice and student growth, including student surveys and measures of student behavior and educational attainment.

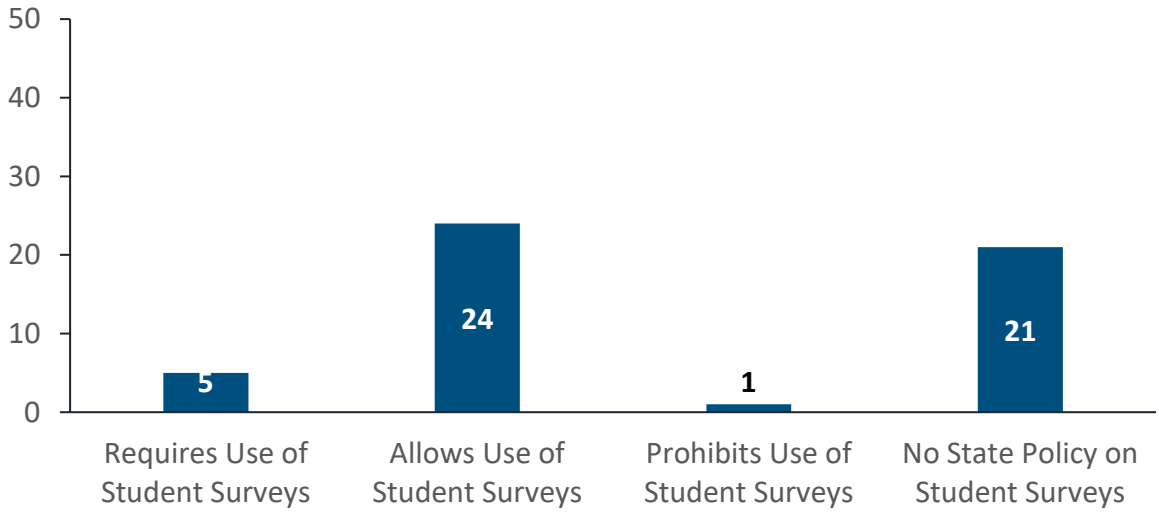
Student Surveys

Student surveys elicit feedback from the most important beneficiaries of effective teaching: students. Well-designed student surveys ask students questions that directly or indirectly assess the effectiveness of their teachers. For example, the Tripod student survey reflects three significant components of effective instruction: content, pedagogy, and relationships (Ferguson & Danielson, 2015). Illinois does not specifically require the use of student surveys in teacher

evaluations at the state level. However, the state stipulates that teacher evaluations should include multiple measures of performance. In their 2016 evaluation of PERA, Milanowski and colleagues reported that 11 of 41 districts within the state chose to use student surveys as an additional measure.

As shown in Exhibit 4, five states require student surveys, and 24 states explicitly allow for them (NCTQ, 2022). States allowing student surveys may see significant variation in how districts implement them. For example, Colorado determined in a statute that using student surveys is a viable measure for evaluating teachers. However, Colorado leaves the decision of using surveys to the district, which may refrain from using them altogether (Close et al., 2020).

Exhibit 4. Number of States That Explicitly Allow, Require, or Prohibit the Use of Student Surveys in Teacher Evaluation, 2022



Note: Data from National Council on Teacher Quality (2022).

Well-designed student surveys can be reliable and valid measures for teacher evaluation. For example, the Measures of Effective Teaching (MET) study (Bill & Melinda Gates Foundation, 2012) found that teachers identified as being in the top 25% of the study sample based on Tripod also tended to have students who made larger test-score gains. Another study found that student surveys are reliable and correlate with measures such as scores from VAMs and classroom observations (Kuhfeld, 2017). While the study doesn't confirm the validity of student surveys for high-stakes use, it makes the case that the instrument could contribute to a summative score along with other measures.

There are concerns that student surveys, when used to measure teacher effectiveness, could be biased due to students' personal feelings toward teachers and by external factors, such as class

size and racial and gender dynamics (Geiger & Amrein-Beardsley, 2017; Marsh, 2007). To address these concerns, surveys have been developed to include indirect questions that gauge instructional quality (Ferguson & Danielson, 2015). Similarly, as some surveys have been administered to children in early grades, including kindergarten, concerns have been raised about the ability of very young children to provide reliable feedback (Geiger & Amrein-Beardsley, 2019). Although there are concerns about the validity of student surveys as a tool for evaluating teacher effectiveness, student surveys do provide a high degree of transparency, as questions are easily understood by evaluated teachers, and the surveys are structured with specific domains that help identify areas for improvement.

Measures of Student Behavior and Educational Attainment

A final group of measures that may be used in teacher evaluation considers teachers' impact on other student outcomes by using data on absences, suspensions, high school graduation, and grade repetition. Research by Jackson (2012; 2018) and Bartanen (2020) suggests that teachers and principals have an impact on student absences and suspensions. Moreover, research by Jackson (2012; 2018) suggests teachers' impact on absences and suspensions seems to be more closely associated with long-term outcomes, such as high school dropout and graduation, SAT completion, and intentions to attend college, than teachers' impact on test-based measures that are widely used for evaluation. Compared to a statistical model that uses only teachers' impact on student test scores to predict student high school graduation and college enrollment, a model that uses both teachers' impact on student test scores and teachers' impact on behavior more than doubles the ability of the model to predict these long-term outcomes (Jackson, 2012; 2018). Therefore, without including nonacademic measures, an evaluation system may overlook the contributions of teachers who excel in a practice that is critically important for their students' long-term outcomes (Jackson, 2012; 2018).

Researchers have used VAM to attempt to isolate and quantify the influence of teachers and leaders on changes in these outcomes while also considering other factors that may affect the outcomes (e.g., Bartanen, 2020; Jackson, 2012, 2018). However, it is important to note that VAMs that attempt to isolate teachers' contributions to student behavior and educational attainment outcomes have not received the same degree of scrutiny as VAMs based on student achievement in ELA and math in Grades 4–8, and that the ASA guidance applies to VAMs based on student behavior or educational attainment outcomes.

State Requirements Regarding District Flexibility in Teacher Evaluation

Over the past decade, teacher evaluation approaches have varied across states. Some impose specific measures, while others allow districts to decide which measures should be used to best evaluate teachers in their districts. Illinois, under PERA, incorporates a mix of general state requirements and local flexibility, an approach also used by 27 other states (NCTQ, 2022). Some

of these 27 states use state standards to define effective teaching, permitting local flexibility in evaluations while ensuring a unified vision (Council of Chief State School Officers, 2013). For example, Colorado allows districts to develop their own evaluation systems in alignment with state standards. These standards cover content knowledge, classroom environment, learning facilitation, and professionalism (Colorado Department of Education, 2023).

Other states that combine state requirements with district flexibility offer optional state models for districts to use. PERA and ISBE guidelines call for implementation of the Illinois state model (which weights the student growth component of a teacher's evaluation at 50%) for cases in which local joint committees cannot reach a consensus on one or more parts of the requirements set forth in the Illinois School Code. State models are commonly used to enhance equity by ensuring districts with limited technical capacity have access to a fully developed state model. For instance, Missouri provides a comprehensive state teacher evaluation model that includes extensive guidance on multiple measures and their application in local teacher evaluation implementation. Districts can use this model and state-provided resources when developing their plans (Missouri Department of Elementary and Secondary Education, 2023).

This approach of combining state requirements with district flexibility contrasts with states adopting a single model to be implemented in all districts (e.g., New Mexico and Tennessee) and those granting complete flexibility to their districts (e.g., Washington, DC). Each approach has benefits. Greater district flexibility can spur innovation and allow systems to better adapt to the local context. Conversely, state-designed teacher evaluation models implemented in all districts can assist in identifying effective or ineffective teachers across the state and highlight disparities across districts in students' access to effective teachers. The effectiveness of better teacher evaluations in improving student outcomes is supported by several studies, with enhancements in student academic performance attributed to structured feedback (Garet et al., 2017), coaching (Allen et al., 2011) and collaboration among teachers (Papay et al., 2016). (See Kane, 2024.)

Using Teacher Evaluation to Support Teacher Development

Teacher evaluations are often intended as assessment tools, but they may also foster teacher growth to enhance student outcomes (Hunter, 2022; Shaha et al., 2015). When evaluation systems offer targeted feedback and pinpoint areas for growth, they may help to improve teachers' professional practices and enrich the educational experience for students (Marzano, 2013b). Illinois is among 30 states that connect teacher evaluations to professional development opportunities or improvement plans; 28 states require that feedback be provided to teachers after classroom observations (Swisher & Saenz-Armstrong, 2022).

Differentiating Teacher Supports. PERA (Illinois General Assembly, 2010) and Illinois Compiled Statutes (Illinois General Assembly, 2024-a) mandate that districts provide teachers rated "Needs Improvement" with a professional development plan within 30 school days. This plan, created in collaboration with the teacher, considers their ongoing professional responsibilities and targets identified improvement areas. It also details any support the district will offer to address these areas. If a tenured teacher receives an "Unsatisfactory" rating, the school district must formulate and initiate a 90-school-day remediation plan within 30 school days to rectify cited deficiencies. Both professional development and remediation plans can extend over 2 school years.[5]

PERA also highlights the importance of additional observations for new teachers to facilitate development and growth. As noted earlier, PERA requires that new, nontenured teachers be observed at least three times each school year, compared to twice every 2 years for more seasoned teachers. PERA requires tenured, experienced teachers who received a "Needs Improvement" or "Unsatisfactory" rating to be observed at least three times in the school year following that rating, regardless of their status.

This emphasis on differentiating teachers and providing frequent observations for early career and low-performing teachers aligns with research suggesting that multiple observations and ongoing feedback improve teacher practice (Connally & Tooley, 2016; Jacob & McGovern, 2015). However, while PERA mandates the creation of professional growth plans for new and struggling teachers, it does not extend this requirement to all teachers. This approach is not unique to Illinois, as many states do not clearly connect their evaluation and support systems (NCTQ, 2022).

Communicating feedback. For each formal observation, PERA requires a preobservation conference between the evaluator and the teacher. During this conference, the teacher is expected to present a written lesson plan or other evidence of planning. The evaluator and teacher discuss the lesson plan or instructional planning and any focus areas for the observation.

After either a formal or informal observation, the evaluator must discuss with the teacher the evidence gathered regarding the teacher's professional practice. If the evidence suggests that the teacher might receive a "Needs Improvement" or "Unsatisfactory" summative performance evaluation rating, the evaluator must inform the teacher.

[5] Since the 2020–2121 school year, the Illinois Compiled Statutes (Illinois General Assembly, 2024) has mandated that districts create and apply a local appeals process for unsatisfactory teacher ratings. This process requires cooperation between the school district and teachers or their representatives. An evaluator panel has the power to reverse "Unsatisfactory" ratings if they are found to be incorrect. Whereas the criteria for appeals are established by a joint committee, the replacements for "unsatisfactory" ratings are decided through negotiation.

Illinois's communication requirements align with the literature on best practices. Teacher evaluations are tools for professional growth (Goe et al., 2012). Clear links between evaluations and development should be maintained, and feedback should be timely, precise, and actionable. Some states encourage using teacher evaluations for growth through online guidance, policy documents, websites, and newsletters. Also, some states have established two-way communication channels, like advisory cabinets, to provide feedback on evaluation initiatives (Connally & Tooley, 2016).

Defining Features of Teacher Evaluation Systems in Illinois, Michigan, Ohio, and Wisconsin

The sections above provide an overview of the Illinois teacher evaluation system established under PERA relative to teacher evaluation systems in the other 49 states and Washington, DC. Exhibit 5 presents a summary of the defining features of teacher evaluation systems of Illinois and three additional states, Michigan, Ohio, and Wisconsin. These states were chosen due to sharing multiple characteristics with Illinois, including region, demographics, and student outcomes. As can be seen, the largest differences between the states are related to the use of student growth measures.

Exhibit 5. Characteristics of State Teacher Evaluation Systems

	Illinois	Michigan	Ohio	Wisconsin
Classroom Observations	State requires all teachers be observed in the classroom	State requires all teachers be observed in the classroom	State requires all teachers be observed in the classroom	State requires all teachers be observed in the classroom
Evaluator Training	State requires teacher evaluators to complete a prequalification process and training	State requires teacher evaluators to be trained, but does not require certification	State requires teacher evaluators to be trained and credentialed	State requires that teacher evaluators receive both training and calibration
Student Growth	State requires test-based student growth measures comprise 30% of teachers' overall ratings	State requires 40% of a teacher's evaluation to be comprised of student growth and assessment data	State requires objective measures of student growth only for teachers of tested grades and subjects; VAM used with state assessment	State requires that 50% of the total evaluation score assigned to a teacher be based upon measures of student performance
SLOs	State allows SLOs to be used	State allows SLOs to be used	State prohibits the use of SLOs	State requires that SLOs be written each year

	Illinois	Michigan	Ohio	Wisconsin
Student Surveys	State policy is silent on student surveys	State policy is silent on student surveys	State policy explicitly allows student surveys	State policy explicitly allows student surveys
District Flexibility	Districts design or adopt their own evaluations system based on criteria from the state	Districts design or adopt their own evaluations system based on criteria from the state	Districts must use the state's evaluation system	Districts may either use the state's evaluation system or develop their own.
Evaluations inform Teacher Professional Development	State requires that evaluations inform professional development for teachers who earn "Unsatisfactory" evaluation ratings	State requires that evaluations inform professional development for all teachers.	State requires that evaluations inform professional development for all teachers	State does not require that evaluations inform professional development

Note: Data from National Council on Teacher Quality (2021–22).

Conclusion

Like many states, Illinois adopted a model that allows district flexibility within the bounds of state-mandated requirements for educator performance evaluation. This approach allows for innovation, diversity, and context-specific application of methods while also offering some comparability across districts.

In line with most states, Illinois includes classroom observations of professional practice as a primary component of its teacher evaluation system, despite the potential for bias and the narrow range of evaluation outcomes that can be associated with this approach. Like several other states, Illinois incorporates SLOs and other student growth measures in its teacher evaluation process. However, unlike other states, Illinois does not mandate the use of its statewide assessments to estimate student growth measures.

PERA includes several best practices on leveraging measures of teacher professional practice as tools for fostering teacher growth. It mandates multiple observations, particularly for new teachers and those requiring improvement. PERA also requires evaluators to hold preobservation conferences with teachers and provide postobservation feedback. While PERA requires the development of professional growth plans for teachers rated "Needs Improvement" or "Unsatisfactory," it does not require this for all teachers. However, research suggests that all teachers can benefit from targeted feedback and ongoing professional development. Ideally, teacher evaluations should promote continuous improvement for all teachers.

As suggested by the policy scan, there is no definitive "right" or "wrong" way to structure teacher evaluations. Each state, including Illinois, must make decisions that inevitably involve trade-offs. The findings in this report emphasize that no single educator evaluation measure is perfectly transparent, valid, and reliable. Thus, a balanced approach, like the one Illinois adheres to under the state's PERA law (Illinois General Assembly, 2010), administrative code (Illinois General Assembly, 2022-a; 2022-b) and nonregulatory guidance (ISBE, 2015), is advantageous as it considers multiple measures and provides diverse perspectives on teachers' practices. Although such flexibility might challenge the comparability of effective educators across different districts, it fosters an evaluation environment that is diverse and contextually appropriate, with each set of choices reflecting its own strengths and challenges.

Review of District Educator Evaluation Plans

As reflected in Research Question (RQ) 1, PEAC and ISBE wanted to know about the scope and substance of the particulars identified in evaluation plans developed by districts and their joint committees. Specifically, information was considered in light of the number of categories that districts used to rate teachers, which types of measures comprise teachers' evaluation ratings, how districts weight measures to create an overall rating, the frequency with which districts evaluate and observe nontenured teachers and tenured teachers based on their previous ratings, whether district evaluation plans include support for addressing teachers in need of improvement, and how and when districts communicate ratings to teachers.

To help understand the components of districts' teacher evaluations, plans were collected and reviewed from a representative sample of Illinois districts. The review was designed to better understand the extent to which plans from across the state incorporated key evaluation components, including components that align with the state's PERA legislation (Illinois General Assembly, 2010), administrative code (Illinois General Assembly, 2022-a; 2022-b), and nonregulatory guidance (ISBE, 2015). The review did *not* include analyses of districts' compliance with state legislation.

Sample Selection

Fifty-one teacher evaluation plans were collected from a representative sample of districts (6% of the 854 districts in the state). Districts were selected using stratified random sampling: districts were sorted into "strata" of similar districts based on their size, locale, and other characteristics. At least two districts from each stratum were included in the analysis to ensure that evaluation plans from a heterogeneous group of districts were studied.

Analysis

The rubric presented in Exhibit 6 was used to review selected district evaluation plans. The team started by coding a small subset of the plans and comparing results between reviewers to ensure consistency in plan evaluation. Once reliability across reviewers was established, the AIR team individually reviewed the remaining plans included in the sample, responding to the rubric questions for the column representing that district; for example, coding "Yes" if there was evidence of the rubric item within the district plan, or coding "Could Not Find Evidence" if there was no evidence of the rubric item within the district plan or if evidence in the evaluation plan was unclear or incomplete.

Limitations

It is important to acknowledge that AIR's review of district evaluation plans does not provide conclusive evidence on the extent to which districts implement their plans as written. For example, although a plan may outline specific measures, such as a requirement that teachers receive three teacher observations per year, the plan does not guarantee that these activities will occur or have been implemented with fidelity. Therefore, the review of these district evaluation plans should be considered only one of multiple data sources.

Exhibit 6. District Evaluation Plan Rubric

Rubric Questions	District A	District B	Etc.
1. Does the plan adhere to or deviate from PERA law?			
a) Does it include a four-tiered rating system?			
b) Are *nontenured* teachers *evaluated* at least once every school year?			
c) Are *tenured* teachers whose performance was last rated either "Needs Improvement" or "Unsatisfactory" *evaluated* at least once every school year?			
d) Are *tenured* teachers whose performance was last rated either "Excellent" or "Proficient" *evaluated* at least once every 3 school years?			
e) Do *nontenured* teachers undergo at least three annual ***observations***?			
f) Do *tenured* teachers whose performance was last rated either "Needs Improvement" or "Unsatisfactory" undergo at least three annual *observations*?			
g) Are all *tenured* teachers whose performance was last rated either "Excellent" or "Proficient" *observed* at least twice in the evaluation cycle?[a]			
2. What measures are included in the evaluation? (checklist)			
a) Test-based student growth measures (including assessment-based Student Learning Objectives)			
b) Non-test-based student growth and/or performance measures (e.g., student portfolios, where teachers guide students in assembling portfolios that showcase their work, achievements, and progress over a specific period)			
c) Teacher performance/professional practice (e.g., planning and preparation; classroom instruction, environment, strategies, and/or behaviors; reflection on teaching; collegiality, professionalism, and professional responsibilities)			

Rubric Questions	District A	District B	Etc.
d) Student surveys			
e) Other measures			
3. For the student growth component, does the evaluation plan require the use of the following?			
a) At least one Type I[b] or Type II[c] assessment			
b) At least one Type III[d] assessment			
4. If the joint committee determines that neither a Type I nor a Type II assessment can be identified, does the evaluation plan require for that at least *two* Type III assessments be used for the student growth component?			
5. Does the plan articulate support and professional development opportunities for teachers to improve their practice? [*Note.* This rubric element refers to support and professional learning opportunities offered to ALL teachers in the district]			
6. Does the plan include professional development for *tenured* teachers whose performance was last rated "Needs Improvement"? [*Note.* This rubric element refers to support and professional learning opportunities offered specifically to teachers who do not meet proficiency thresholds.]			
7. Does the plan include a process for remediation for tenured teachers whose performance was last rated "Unsatisfactory"? [*Note.* This rubric element refers to support and professional learning opportunities offered specifically to teachers who do not meet proficiency thresholds.]			
8. Does the plan articulate how the various components should be combined to yield an overall rating for an educator?			
9. How much are each of the following components weighted (0%–100%) in teachers' evaluations?			
• Test-based student growth measures			
• Non-test-based student growth or performance measures			
• Teacher performance/professional practice			

Rubric Questions	District A	District B	Etc.
• Student surveys			
• Other measures			
10. Does the plan articulate the **date** by which ratings must be communicated to educators?			
11. Does the plan articulate **how** ratings should be communicated to educators?			

[a] *Evaluation Cycle* refers to the frequency at which the teacher is evaluated, which is every other year (at minimum) for tenured teachers receiving "Excellent" or "Proficient" ratings, every year for nontenured teachers, and every year for tenured teachers receiving "Needs Improvement" or "Unsatisfactory" ratings.
[b] *Type I assessment* means a reliable assessment that measures a certain group or subset of students in the same manner with the same potential assessment items, is scored by a nondistrict entity, and is administered either statewide or beyond Illinois. Examples include assessments available from the Northwest Evaluation Association (NWEA), Scantron Performance Series, Star Reading Enterprise, College Board's SAT, Advanced Placement, or International Baccalaureate examinations, or ACT's EPAS® (i.e., Educational Planning and Assessment System) (Illinois State Board of Education [ISBE], 2015).
[c] *Type II assessment* means any assessment developed or adopted and approved for use by the school district and intended to be used on a districtwide basis by all teachers in a given grade, course, or subject area. Examples include collaboratively developed common assessments, curriculum tests and assessments designed by textbook publishers (ISBE, 2015).
[d] *Type III assessment* means any assessment that is rigorous, that is aligned to the course's curriculum, and that the qualified evaluator and teacher determine measures student learning in that course. Examples include teacher-created assessments, assessments designed by textbook publishers, student work samples or portfolios, assessments of student performance, and assessments designed by staff who are subject or grade-level experts that are administered commonly across a given grade or subject. A Type I or Type II assessment may qualify as a Type III assessment if it aligns with the curriculum being taught and measures student learning in that subject area (ISBE, 2015).

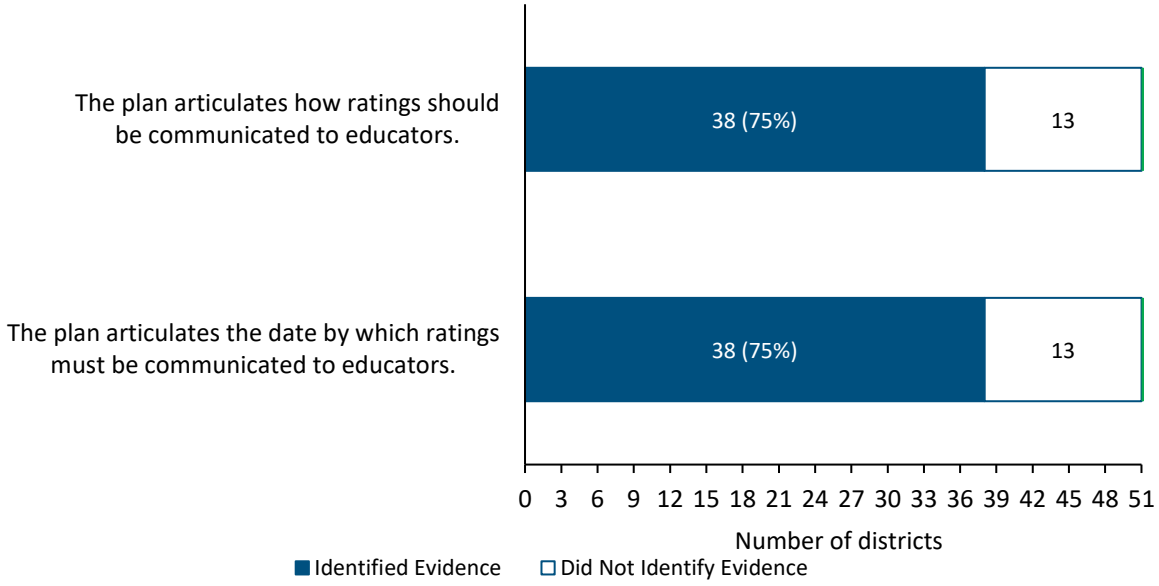
Findings

In this section, findings from the review of 51 district teacher evaluation plans are shared.

Teacher Evaluation Rating Categories and Communication

The review of districts' teacher evaluation plans included an analysis to identify the number of categories that districts use to rate teachers. All 51 evaluation plans selected for review included four categories for rating teachers. The review of district evaluation plans also included how and when districts communicate ratings to their teachers. Exhibit 7 shows that 35 of 51 evaluation plans provided evidence of how districts communicate ratings to educators. In addition, 30 of 51 evaluation plans included evidence of the date by which districts must communicate ratings to educators.

Exhibit 7. Number of District Evaluation Plans With Evidence of How and When Ratings Should Be Communicated to Educators



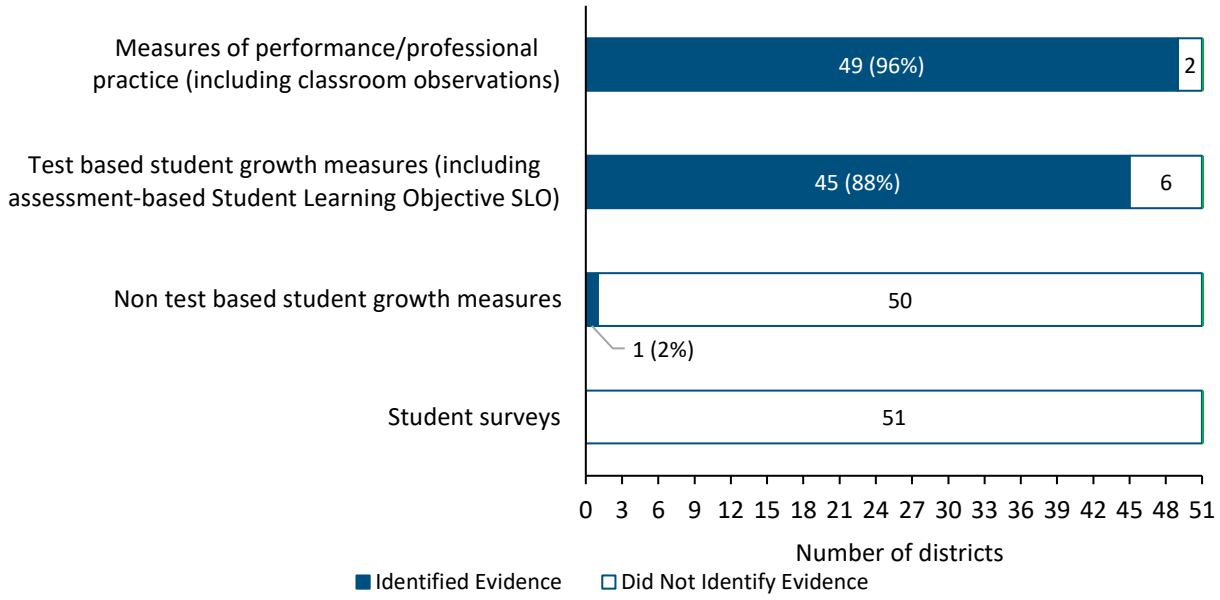
Measures Included in Teacher Evaluation Ratings

District evaluation plans were evaluated for the types of measures that contribute to teachers' evaluation ratings, including the specific types of student assessments included in these ratings, and how districts weighted each measure to yield an overall teacher evaluation rating.

Exhibit 8 shows the number of teacher evaluation plans that reported the use of teacher performance or professional practice measures, test-based and non-test-based student growth measures, and student surveys in teacher ratings. Forty-nine of the 51 district evaluation plans used educator performance or professional practice measures, and 45 plans included student

growth measures. One district plan included the use of non-test-based student growth measures, and no district plan included the use of student surveys.

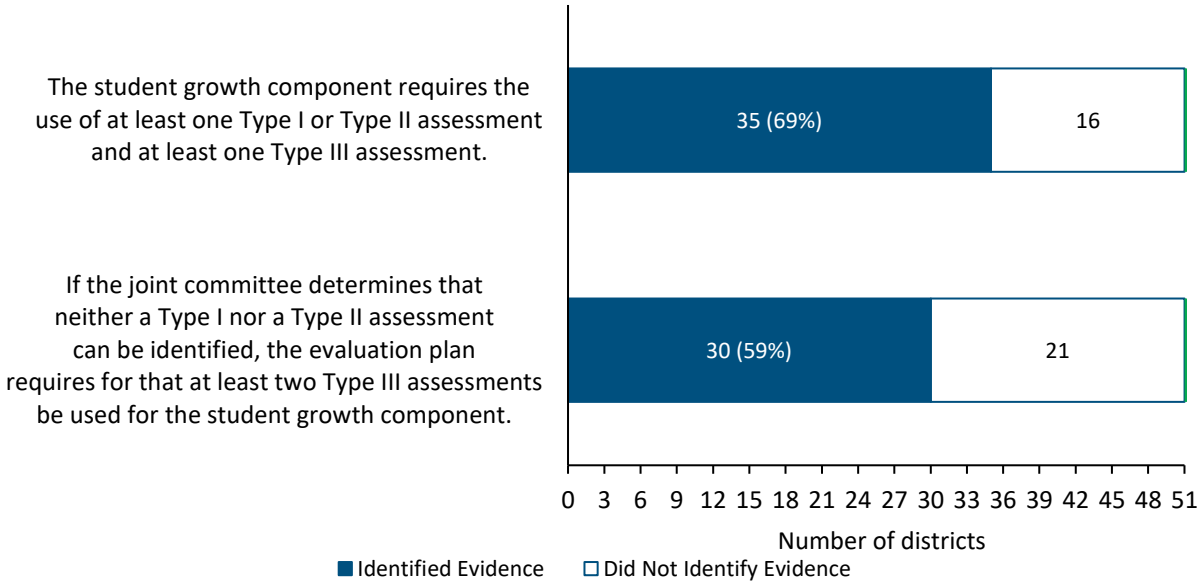
Exhibit 8. Number of District Evaluation Plans With Evidence of Measures Included in Their Evaluation System



In 36 of 51 evaluation plans, the professional practice measure comprised 70% of teachers' ratings, and the test-based student growth measure comprised 30% of teachers' evaluation ratings. One district weighed both test-based student growth and teacher performance at 50% each. The remaining districts did not report specific weights of the two required components.

Exhibit 9 shows that 35 of 51 districts include at least one Type I or Type II assessment and at least one Type III assessment in their evaluation plans. In addition, 30 districts also include at least one Type III assessment.

Exhibit 9. Number of District Evaluation Plans With Evidence of the Types of Student Assessments



Frequency of Teacher Evaluations and Observations

The frequency of teacher evaluations, observations by teachers' tenure status, and observations by teachers' prior performance rating varied across the sample of 51 districts. Specifically, the review of evaluation plans included identifying the extent to which districts observe and evaluate the following three categories of teachers: (a) tenured teachers whose last performance rating placed them in the higher performance categories (i.e., "Excellent" or "Proficient"), (b) tenured teachers whose last performance rating placed them in the lowest performance categories (i.e., "Needs Improvement" or "Proficient"), and (c) nontenured teachers.

Exhibit 10 shows the number of reviewed district evaluation plans that included evidence that districts evaluate tenured teachers who were rated "Excellent" or Proficient" *at least every 3 school years* and that districts evaluate tenured teachers who were rated "Needs Improvement" or "Unsatisfactory" and nontenured teachers *every school year*. Forty-three of the 51 evaluation plans included evidence that districts evaluate tenured teachers with favorable ratings at least once every 3 years and districts evaluate nontenured teachers at least once per school year. In addition, 40 of the 51 plans showed evidence that districts evaluate tenured teachers with unfavorable ratings at least once per school year.

Exhibit 10. Number of District Evaluation Plans With Evidence of Evaluation Frequencies for Tenured and Nontenured Teachers

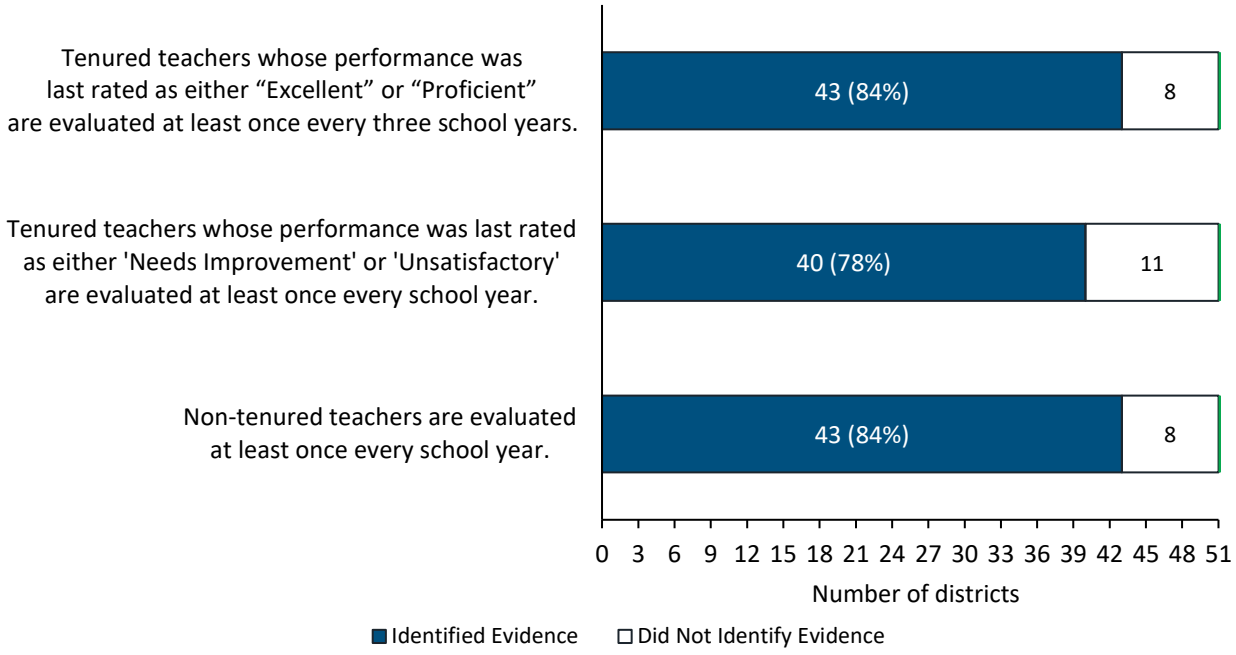
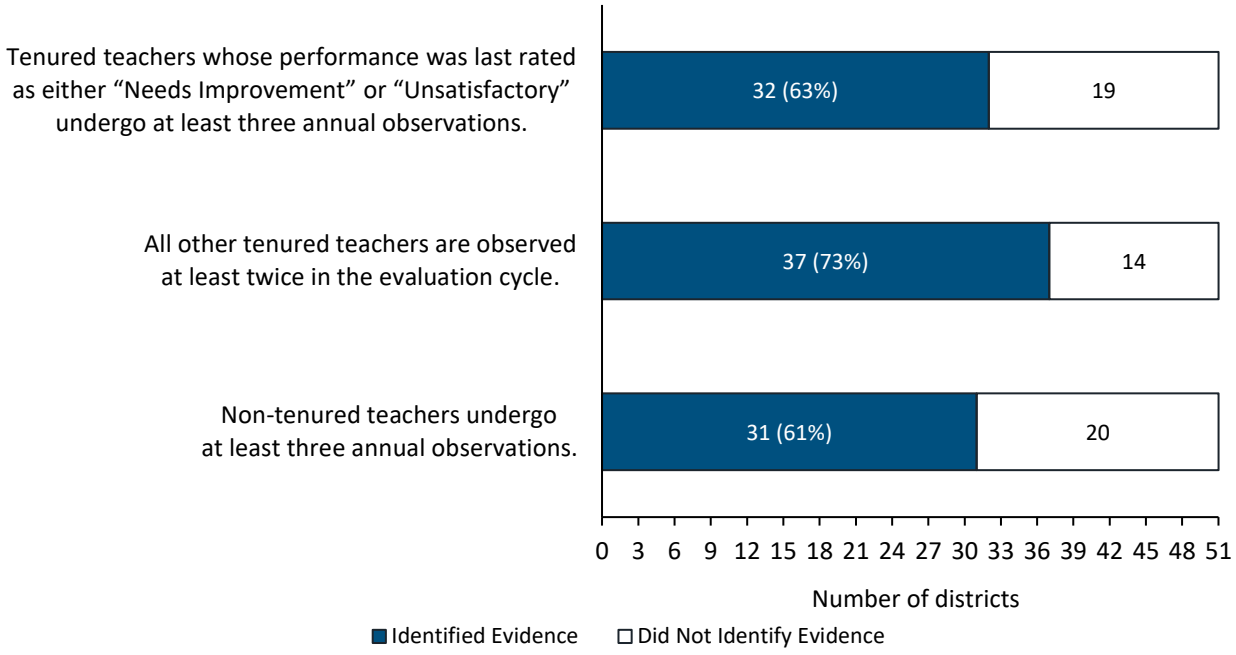


Exhibit 11 shows the number of reviewed district evaluation plans that included evidence that districts *observe* tenured teachers who were rated "Needs Improvement" or "Unsatisfactory" and nontenured teachers at least three times every school year and observe all other teachers at least twice during their evaluation cycle. Thirty-two out of 51 evaluation plans had evidence that districts observe tenured teachers who were rated "Needs Improvement" or "Unsatisfactory" at least three times annually, while 37 districts observe all other teachers at least twice during the evaluation cycle. Nontenured teachers receive three annual observations, according to 31 evaluation plans.

Exhibit 11. Number of District Evaluation Plans With Evidence of Observation Frequencies for Tenured and Nontenured Teachers



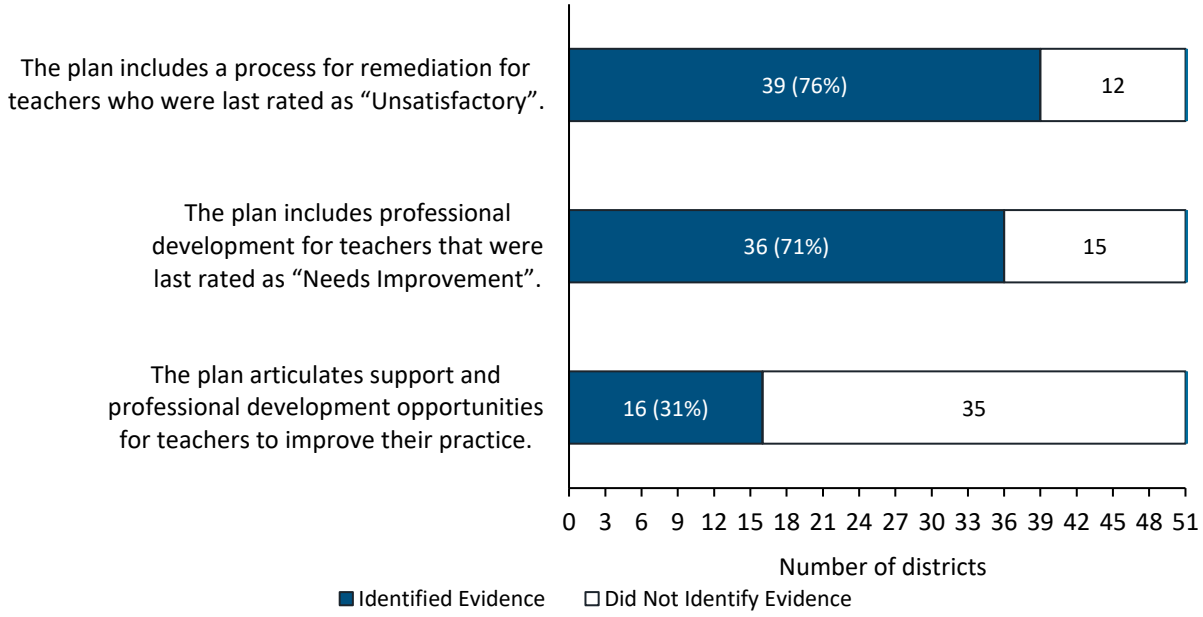
Support and Professional Learning Opportunities for Teachers

The sample of districts' teacher evaluation plans was reviewed to understand whether districts provide support to teachers to address areas identified as needing improvement as a result of their evaluation. Specifically, plans were reviewed to understand whether they include remediation for tenured teachers whose performance was last rated "Unsatisfactory" and professional development for tenured teachers whose performance was last rated "Needs Improvement, as well as whether districts' teacher evaluation plans include support and professional development opportunities for all teachers to improve their practice.

Exhibit 12 shows that of the 51 district evaluation plans, 36 prescribe professional development plans for teachers rated "Needs Improvement," and 39 prescribe remediation plans for teachers rated "Unsatisfactory."[6] Moreover, 16 of 51 district evaluation plans had evidence of using the evaluation process to inform professional development for all teachers.

[6] Although some district plans did not explicitly state that professional development plans are required for teachers rated "Needs Improvement" and/or remediation plans are required for teachers rated "Unsatisfactory," this is not meant to imply that these districts do not implement these requirements.

Exhibit 12. Number of District Evaluation Plans With Evidence of Professional Development Opportunities for Teachers



Summary

Nine members of the AIR team used a rubric to review district evaluation plans. They responded to the rubric questions for the column representing each district, coding "Yes" if there was evidence of the rubric item within the district's plan, or coding "Could Not Find Evidence" if the item was missing or if evidence in the evaluation plan was unclear or incomplete. The team began this process by coding a small subset of the plans and comparing results between reviewers to guarantee consistency in plan evaluation. Once they had established reliability across reviewers, the AIR team individually reviewed the remaining plans included in the sample.

In sum, the review identified the following:

- All 51 evaluation plans rated teachers on a four-category rating system: "Excellent," "Proficient," "Needs Improvement," and "Unsatisfactory."
- Almost all districts reported the use of performance or professional practice measures and test-based student growth measures in their ratings of teachers.
- In 36 of 51 district evaluation plans, performance or professional practice measures comprised 70% and test-based student growth measures comprised 30% of teachers' overall ratings.

- In 43 evaluation plans, nontenured teachers are evaluated every year, and tenured teachers whose performance was last rated "Excellent" or "Proficient" are evaluated every 3 years. In 40 evaluation plans, teachers whose performance was last rated "Needs Improvement" or "Unsatisfactory" are evaluated every year.
- in 32 plans, tenured teachers who were rated "Needs Improvement" or "Unsatisfactory" are observed at least three times; in 31 plans, nontenured teachers are observed at least three times annually; and in 37 plans, teachers are observed at least twice during the evaluation cycle.
- Thirty-six districts require professional development plans for teachers rated "Needs Improvement," and 39 districts require remedial plans for teachers rated "Unsatisfactory."
- In less than a third of the plans (16 out of 51), districts went beyond the required professional development for teachers who received "Unsatisfactory" and "Needs Improvement" ratings, and described how they would use the evaluation to inform professional development for all teachers.
- There was no evidence that districts include measures beyond professional practice and student growth in their teacher evaluation plans.

Educator Evaluation Ratings Patterns and Trends

RQ6 asks, "What are the distributions of educator ratings across districts and schools and over time, and to what extent do educator ratings vary systematically with respect to the characteristics of educators, districts, and schools?"

To answer this question, a set of secondary RQs was developed:

- RQ6a. What percentage of educators received evaluation ratings in each of the four categories ("Excellent," "Proficient," "Needs Improvement," and "Unsatisfactory")?
- RQ6b. What percentage of educators received ratings in each of the categories by teacher, school, and district characteristics?
- RQ6c. How did the distributions of evaluation ratings change over time?

Data Sources

Data used to answer these RQs include ISBE educator-level data for 2013–14 to 2022–23 and Illinois Report Card data for 2022–23 and 2021–22. Exhibit 13 shows the data sources and data elements used in the analyses.

Exhibit 13. Data Sources and Elements

Data Sources	Data Elements
ISBE educator-level data, 2013–14 to 2022–23	• Educator evaluation ratings • Educator characteristics: gender, race/ethnicity, years of experience in district, work role
The Illinois Report Card 2022–23	• School characteristics: school level, locale, low-income student percentage • District characteristics: low-income student percentage, percentage of teachers who are White, average teacher years of experience, teacher retention rate
The Illinois Report Card 2021–22	• School characteristics: student ELA and math achievement proficiency percentages • District characteristics: student ELA and math achievement proficiency percentages

Analysis and Findings

RQ6a. What percentage of educators received evaluation ratings in each of the four categories ("Excellent," "Proficient," "Needs Improvement," and "Unsatisfactory")?

ISBE educator-level data for 2022–23 were used to answer secondary RQ6a and RQ6b. This data set included data for 145,663 teachers, principals, and assistant principals in 4,364 schools in 1,000 districts. Of these educators, 141,031 worked in only one school, and 4,632 worked in

two or more schools or districts. Educators in this analysis included those who received only one evaluation rating across all schools or districts where they worked. Because the data sometimes included multiple records for the same educator, only one record per educator was retained.

To answer RQ6a, the percentage of educators who received a rating in each of the categories separately for teachers and principals/assistant principals was calculated. If the 2022–23 data did not include a rating for a given teacher, principal, and assistant principal, that educator was assigned to a fifth rating category, "Missing Rating." The ISBE data team noted that there was no explanation for why many educators were missing ratings. In the exhibits below, the percentages of educators who received each category of ratings. The top panel of each exhibit shows all educators, including those who were "Missing Rating." The bottom panel shows only educators who received a rating. Groups with any cell size of fewer than 10 educators are suppressed to protect educator privacy (U.S. Department of Education, 2023; Seastrom, 2010).

The results for teacher evaluation ratings are shown in Exhibit 14. Of the 138,639 teachers in the data, 62.6% did not receive an evaluation rating in 2022–23. Of the teachers who received a rating, 97.2% were rated "Excellent" or "Proficient." Only 2.6% of the evaluated teachers received a "Needs Improvement" rating and 0.2% were rated "Unsatisfactory." Because percentages in the evaluation categories of "Unsatisfactory" and "Needs Improvement" are very low, the two groups and report combined percentages. This approach is used for the remainder of this section to protect educator privacy.

Exhibit 14. 2022–23 Teacher Evaluation Ratings

	Number of Teachers	Excellent	Proficient	Needs Improvement	Unsatisfactory	Missing Rating
Including Missing	138,639	17.3%	19.0%	1.0%	0.1%	62.6%
Excluding Missing	51,843	46.4%	50.8%	2.6%	0.2%	-

The distribution of principal and assistant principal evaluation ratings is very similar to teacher ratings (Exhibit 15). Over two thirds (68.0%) of principals and assistant principals did not receive an evaluation rating in 2022–23. Of evaluated principals and assistant principals, only 3.2% received "Needs Improvement" or "Unsatisfactory" ratings. The other 96.8% were rated either "Excellent" or "Proficient."

Exhibit 15. 2022–23 Principal/Assistant Principal Evaluation Ratings

	Number of Principals/APs	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing	7,024	15.0%	16.0%	1.0%	68.0%
Excluding Missing	2,246	46.9%	49.9%	3.2%	-

Of educators who received an evaluation rating in 2022–23, 97.2% of teachers and 96.8% of principals and assistant principals received a rating of "Excellent" or "Proficient."

RQ6b. What percentage of educators received ratings in each of the categories by teacher, school, and district characteristics?

To answer RQ6b, the percentage of educator ratings in each of the rating categories by educator, school, and district characteristics was calculated for educators who received only one evaluation rating across all schools or districts where they worked. For analyses that examined evaluation ratings by school- or district-level characteristics, educators who worked at two or more schools or districts with different values for each characteristic were coded as having "Multiple Values." Educators who worked at a prekindergarten school in addition to an elementary, middle, or high school were assigned the school and district characteristics of the non-prekindergarten school. A total of 3,767 educators in 535 schools were not matched to the Illinois Report Card data. For educators not matched to any school in the Illinois Report Card data, school- and district-level characteristics were coded as "Missing."

In addition, some categories with small cell sizes were combined to ensure that all groups with reported results had 10 or more members to protect educator privacy. Examples of cases where categories were combined to protect educator privacy include recoding the race/ethnicity categories "American Indian or Alaska Native," "Asian," "Native Hawaiian or Other Pacific Islander," "Two or More Races," and "Unknown" as "Other"; combining the school and district "Multiple Values" and "Missing" categories; and combining the evaluation rating categories "Unsatisfactory" and "Needs Improvement."

Teacher Evaluation Ratings

The distribution of evaluation ratings varied by teachers' gender, race/ethnicity, and years of experience. Exhibit 16 shows the evaluation ratings for teachers by gender. In 2022–23, 63.3% of male teachers and 62.4% of female teachers did not receive a rating. Of the teachers who were evaluated, male teachers had lower percentages than female teachers in the rating

categories of “Excellent” (41.8% versus 47.7%) and higher percentages in “Proficient” and “Needs Improvement” or “Unsatisfactory.”

Exhibit 16. 2022–23 Teacher Evaluation Ratings by Teacher Gender

Teacher Gender	Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
Male	32,508	15.4%	20.0%	1.4%	63.3%
Female	106,131	18.0%	18.7%	0.9%	62.4%
Excluding Missing					
Male	11,934	41.8%	54.4%	3.8%	-
Female	39,909	47.7%	49.7%	2.5%	-

Note. Percentages and total number of teachers apply to each row.

Exhibit 17 shows teacher evaluation ratings by race/ethnicity. Black or African American teachers had a much higher rate for missing ratings (71.0%) in 2022–23 than other groups (e.g., 58.5% for White teachers). Compared to White teachers who received a rating, a lower percentage of Black or African American teachers, Hispanic or Latinx teachers, and teachers of other races/ethnicities who received a rating were rated “Excellent,” and a higher percentage were rated “Needs Improvement” or “Unsatisfactory.” Notably, only 22.1% of Black or African American teachers were rated “Excellent” compared to 50.0% of White teachers, and 10.6% of Black or African American teachers were rated “Needs Improvement” or “Unsatisfactory” compared to only 2.0% of White teachers.

Exhibit 17. 2022–23 Teacher Evaluation Ratings by Teacher Race/Ethnicity

Teacher Race/Ethnicity	Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
Black or African American	8,745	6.4%	19.5%	3.1%	71.0%
Hispanic or Latinx	11,676	12.4%	24.2%	2.1%	61.2%
White	111,508	18.8%	18.1%	0.7%	62.3%
Other	6,710	15.2%	24.4%	1.8%	58.5%

Teacher Race/Ethnicity	Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Excluding Missing					
Black or African American	2,533	22.1%	67.3%	10.6%	-
Hispanic or Latinx	4,533	32.0%	62.5%	5.5%	-
White	41,995	50.0%	48.0%	2.0%	-
Other	2,782	36.7%	58.9%	4.4%	-

Note. Percentages and total number of teachers apply to each row.

The distribution of teacher ratings in 2022–23 differed across teachers' years of experience in their districts. As shown in Exhibit 18, 60.6% of first-year teachers were evaluated in 2022–23 compared to only 25.5% of teachers with 6 or more years of experience. Of teachers who were evaluated, the percentage receiving "Needs Improvement" or "Unsatisfactory" was the highest for first-year teachers (5.7%) and lower with each category of additional experience (e.g., only 1.2% of teachers with 6 or more years of experience were rated "Needs Improvement" or "Unsatisfactory").

Exhibit 18. 2022–23 Teacher Evaluation Ratings by Teacher Years of Experience

Teacher Years of Experience	Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
1st Year	15,720	13.1%	44.0%	3.5%	39.4%
2 or 3 Years	19,037	24.1%	42.1%	2.6%	31.1%
4 or 5 Years	15,644	19.5%	22.4%	1.0%	57.1%
6 or More Years	88,238	16.2%	9.0%	0.3%	74.5%
Excluding Missing					
1st Year	9,521	21.7%	72.6%	5.7%	-
2nd or 3rd Year	13,113	35.0%	61.1%	3.8%	-
4th or 5th Year	6,714	45.5%	52.2%	2.3%	-
6th or More Year	22,495	63.7%	35.1%	1.2%	-

Note. Percentages and total number of teachers apply to each row.

To examine variations by race within teaching experience categories, AIR disaggregated teacher evaluation ratings data, as shown in Exhibit 19 to Exhibit 22. Among teachers who were evaluated in 2022–23 within each experience category, Black or African American and Hispanic or Latinx teachers received "Excellent" ratings at much lower rates and "Needs Improvement" or "Unsatisfactory" ratings at much higher rates than White teachers. For example, among first-year teachers, 10.8% of Black or African American teachers and 15.0% of Hispanic or Latinx teachers were rated as "Excellent", compared with 24.0% of White teachers; 12.6% of Black or African American teachers and 9.3% of Hispanic or Latinx teachers were rated as "Needs Improvement" or "Unsatisfactory", compared with 4.6% of White teachers (Exhibit 20). Results by race/ethnicity for teachers with 2 to 3 years of experience, teachers with 4 to 5 years of experience, and teachers with 6 or more years of experience are similar; that is, Black or African American teachers and Hispanic or Latinx teachers received lower ratings than White teachers with similar years of teaching experience.

Exhibit 19. First-Year Teacher Evaluation Ratings by Race/Ethnicity

Teacher Race/Ethnicity	Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
Black or African American	1,361	4.4%	31.4%	5.1%	59.1%
Hispanic or Latinx	1,623	9.5%	47.9%	5.9%	36.7%
White	11,489	15.2%	45.2%	2.9%	36.7%
Other	1,247	8.7%	41.3%	3.5%	46.5%
Excluding Missing					
Black or African American	557	10.8%	76.7%	12.6%	-
Hispanic or Latinx	1,027	15.0%	75.7%	9.3%	-
White	7,270	24.0%	71.4%	4.6%	-
Other	667	16.2%	77.2%	6.6%	-

Note. Percentages and total number of teachers apply to each row.

Exhibit 20. Teachers With 2 or 3 Years of Experience, Evaluation Ratings by Race/Ethnicity

Teacher Race/Ethnicity	Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
Black or African American	1,469	9.1%	40.6%	5.4%	44.8%
Hispanic or Latinx	2,014	20.7%	48.4%	4.8%	26.1%
White	14,243	26.6%	41.1%	1.9%	30.4%
Other	1,311	19.4%	45.7%	3.7%	31.2%
Excluding Missing					
Black or African American	811	16.5%	73.6%	9.9%	-
Hispanic or Latinx	1,488	28.0%	65.5%	6.5%	-
White	9,912	38.2%	59.0%	2.8%	-
Other	902	28.2%	66.4%	5.4%	-

Note. Percentages and total number of teachers apply to each row.

Exhibit 21. Teachers With 4 or 5 Years of Experience, Evaluation Ratings by Race/Ethnicity

Teacher Race/Ethnicity	Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
Black or African American	926	10.3%	23.5%	3.8%	62.4%
Hispanic or Latinx	1,397	16.1%	29.3%	1.9%	52.6%
White	12,396	20.5%	21.6%	0.6%	57.3%
Other	925	21.2%	21.8%	1.3%	55.7%
Excluding Missing					
Black or African American	348	27.3%	62.6%	10.1%	-

Teacher Race/Ethnicity	Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Hispanic or Latinx	662	34.0%	61.9%	4.1%	-
White	5,294	48.0%	50.5%	1.5%	-
Other	410	47.8%	49.3%	2.9%	-

Note. Percentages and total number of teachers apply to each row.

Exhibit 22. Teachers With 6 or More Years of Experience, Evaluation Ratings by Race/Ethnicity

Teacher Race/Ethnicity	Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
Black or African American	4,989	5.5%	9.3%	1.7%	83.6%
Hispanic or Latinx	6,642	9.9%	10.1%	0.5%	79.6%
White	73,380	17.6%	8.8%	0.2%	73.4%
Other	3,227	14.3%	10.0%	0.5%	75.1%
Excluding Missing					
Black or African American	817	33.3%	56.5%	10.2%	-
Hispanic or Latinx	1,356	48.4%	49.4%	2.2%	-
White	19,519	66.3%	33.0%	0.7%	-
Other	803	57.7%	40.2%	2.1%	-

Note. Percentages and total number of teachers apply to each row.

The distribution of teacher evaluation ratings also differed by school characteristics, such as Regional Office of Education (ROE) Service Area,[7] locale, school level, and low-income student enrollment. Exhibit 23 shows the percentage of teachers who received each rating by their school's ROE service area. Teachers in Service Area 2 had the highest rate of being evaluated in 2022–23, with only 54.4% missing data compared to 61.7% or higher for other service areas.

[7] See Illinois Association of Regional School Superintendents (2015).

Among teachers who were evaluated, those in ROE Service Areas 1 and 3 had the lowest percentages for receiving an "Excellent" rating and the highest percentages for receiving "Needs Improvement" or "Unsatisfactory."

Exhibit 23. 2022–23 Teacher Evaluation Ratings by School ROE Service Area

ROE Service Area	Total Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
1	91,036	16.1%	19.1%	1.3%	63.5%
2	11,003	21.2%	23.9%	0.5%	54.4%
3	9,387	16.1%	20.5%	1.1%	62.4%
4	9,684	16.9%	17.8%	0.4%	65.0%
5	8,118	24.6%	13.3%	0.4%	61.7%
6	5,776	21.1%	14.4%	0.4%	64.2%
Multiple or Missing ROE Service Area	3,635	19.2%	20.6%	1.4%	58.8%
Excluding Missing					
1	33,223	44.1%	52.4%	3.5%	-
2	5,022	46.5%	52.3%	1.2%	-
3	3,534	42.7%	54.5%	2.8%	-
4	3,392	48.1%	50.7%	1.2%	-
5	3,106	64.2%	34.8%	0.9%	-
6	2,069	58.8%	40.1%	1.2%	-
Multiple or Missing ROE Service Area	1,497	46.6%	50.0%	3.4%	-

Note. Percentages and total number of teachers apply to each row.

Ratings also differed by school locale (Exhibit 24). Of teachers who were evaluated, the percentage of city teachers receiving an "Excellent" rating was much lower (32.8%) than in other locales (49.9% or more), and their percentage of receiving "Needs Improvement" or "Unsatisfactory" ratings was much higher (7.0% versus 1.7% or lower) than in other locales.

Exhibit 24. 2022–23 Teacher Evaluation Ratings by School Locale

School Locale	Total Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
City	37,475	10.3%	18.8%	2.2%	68.7%
Suburb	66,901	20.5%	20.0%	0.6%	58.8%
Town	13,712	18.8%	17.3%	0.5%	63.4%
Rural	16,038	18.9%	16.7%	0.6%	63.7%
Multiple or Missing Locale	4,513	18.5%	19.0%	1.3%	61.2%
Excluding Missing					
City	11,715	32.8%	60.2%	7.0%	-
Suburb	27,541	49.9%	48.6%	1.5%	-
Town	5,022	51.4%	47.3%	1.3%	-
Rural	5,814	52.3%	46.0%	1.7%	-
Multiple or Missing Locale	1,751	47.6%	48.9%	3.4%	-

Note. Percentages and total number of teachers apply to each row.

Exhibit 25 shows the distribution of ratings by school level.[8] Among teachers who received a rating, higher percentages of teachers in middle or high school received higher ratings than teachers in elementary schools. Teachers who taught in elementary schools had the lowest percentage in the "Excellent" category (43.5% compared to 44.3% for middle school and 52.7% for high school) and the highest percentage in the "Needs Improvement" or "Unsatisfactory" category (3.6% versus 1.7% and 2.1%, respectively).

[8] School-level categories from the ISBE Illinois Report Card 2022–23 were used.

Exhibit 25. 2022–23 Teacher Evaluation Ratings by School Level

School Level	Total Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
PreK	1067	S	S	S	S
Elementary	70145	16.0%	19.5%	1.3%	63.2%
Middle	21424	17.8%	21.7%	0.7%	59.8%
High	40476	19.1%	16.3%	0.8%	63.8%
Multiple or Missing School Level	5527	19.4%	20.2%	1.3%	59.1%
Excluding Missing					
PreK	488	S	S	S	-
Elementary	25836	43.5%	53.0%	3.6%	-
Middle	8620	44.3%	53.9%	1.7%	-
High	14639	52.7%	45.2%	2.1%	-
Multiple or Missing School Level	2260	47.5%	49.4%	3.1%	-

Note. Percentages and total number of teachers apply to each row. *S* indicates that data were suppressed because at least one cell in the row represents data for fewer than 10 educators.

To examine the potential relationship between schools' student demographics and teacher ratings, cross tabulations for teacher ratings and each school's low-income student enrollment percentage quartile and prior-year (2021–22) ELA and math achievement proficiency rate quartiles were conducted. As shown in Exhibit 26, among teachers who were evaluated, those at schools with higher percentages of low-income[9] students had higher percentages of receiving "Needs Improvement" or "Unsatisfactory" ratings (8.3% for Quartile 4 versus 1.2% for Quartile 1) and lower percentages of receiving an "Excellent" rating (26.3% for Quartile 4 versus 55.4% for Quartile 1) compared to teachers at schools with lower percentages of low-income students. The results for prior-year ELA and math achievement quartiles (Exhibit 27 and

[9] The ISBE Illinois Report Card 2022–23 defines low-income students as "those who receive or live in households that receive Supplemental Nutrition Assistance Program or Temporary Assistance to Needy Families benefits; are classified as homeless, migrant, runaway, Head Start, or foster children; or live in a household where the household income meets the U.S. Department of Agriculture income guidelines to receive free or reduced-price meals."

Exhibit 28, respectively) similarly demonstrated that a higher percentage of teachers working in disadvantaged school contexts received lower evaluation ratings.

Exhibit 26. 2022–23 Teacher Evaluation Ratings by School Low-Income Student Enrollment Quartile

School Low-Income Student Enrollment Quartile	Total Number of Teacher	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
1st (Percentile 1–25)	40,621	21.8%	17.0%	0.5%	60.7%
2nd (Percentile 26–50)	29,774	18.9%	18.2%	0.6%	62.3%
3rd (Percentile 51–75)	32,219	17.6%	21.5%	0.8%	60.1%
4th (Percentile 76–99)	29,049	8.1%	20.1%	2.6%	69.2%
Multiple or Missing Quartile	6,976	22.1%	17.7%	1.1%	59.1%
Excluding Missing					
1st (Percentile 1–25)	15,958	55.4%	43.4%	1.2%	-
2nd (Percentile 26–50)	11,233	50.2%	48.3%	1.6%	-
3rd (Percentile 51–75)	12,863	44.1%	53.8%	2.1%	-
4th (Percentile 76–99)	8,935	26.3%	65.4%	8.3%	-
Multiple or Missing Quartile	2,854	54.0%	43.3%	2.7%	-

Note. Percentages and total number of teachers apply to each row.

Exhibit 27. 2022–23 Teacher Evaluation Ratings by 2021–22 School ELA Proficiency Rate Quartile

School ELA Proficiency Rate Quartile	Total Number of Teacher	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
1st (Percentile 1–25)	29,097	11.2%	23.3%	2.2%	63.3%
2nd (Percentile 26–50)	27,613	18.0%	19.8%	1.2%	61.0%
3rd (Percentile 51–75)	28,495	18.3%	17.5%	0.6%	63.7%
4th (Percentile 76–99)	35,157	21.8%	17.2%	0.5%	60.5%

School ELA Proficiency Rate Quartile	Total Number of Teacher	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Multiple or Missing Quartile	18,277	16.2%	16.7%	0.8%	66.3%
Excluding Missing					
1st (Percentile 1–25)	10,684	30.4%	63.6%	6.0%	-
2nd (Percentile 26–50)	10,765	46.1%	50.8%	3.0%	-
3rd (Percentile 51–75)	10,356	50.3%	48.1%	1.7%	-
4th (Percentile 76–99)	13,882	55.2%	43.5%	1.3%	-
Multiple or Missing Quartile	6,156	48.1%	49.7%	2.3%	-

Note. Percentages and total number of teachers apply to each row.

Exhibit 28. 2022–23 Teacher Evaluation Ratings by 2021–22 School Math Proficiency Rate Quartile

School Math Proficiency Rate Quartile	Total Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
1st (Percentile 1–25)	27,994	10.7%	23.9%	2.3%	63.2%
2nd (Percentile 26–50)	26,817	17.0%	19.1%	1.2%	62.7%
3rd (Percentile 51–75)	27,846	18.6%	18.1%	0.7%	62.6%
4th (Percentile 76–99)	38,083	22.2%	17.1%	0.5%	60.2%
Multiple or Missing Quartile	17,899	16.1%	16.7%	0.7%	66.5%
Excluding Missing					
1st (Percentile 1–25)	10,296	29.0%	64.9%	6.1%	-
2nd (Percentile 26–50)	9,999	45.7%	51.2%	3.1%	-
3rd (Percentile 51–75)	10,420	49.7%	48.5%	1.9%	-
4th (Percentile 76–99)	15,139	55.7%	43.0%	1.3%	-
Multiple or Missing Quartile	5,989	48.0%	49.8%	2.2%	-

Note. Percentages and total number of teachers apply to each row.

AIR disaggregated teacher evaluation data for the two highest quartiles of low-income student enrollment and report the percentage of ratings each teacher racial/ethnic group received in (Exhibit 29 and Exhibit 30). Within each quartile of low-income student enrollment, a lower proportion of Black or African American teachers received "Excellent" ratings than White teachers (31.3% vs. 46.8% for 3rd quartile; 13.3% vs. 31.4% for 4th quartile). Because some cell sizes are fewer than 10 teachers, to protect teacher privacy, evaluation ratings data are not reported for the two lowest quartiles of low-income student enrollment by teacher race/ethnicity.

Exhibit 29. 2022–23 Teacher Evaluation Ratings in Schools in the 3rd Quartile of Low-Income Enrollment by Teacher Race/Ethnicity

Teacher Race/Ethnicity	Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
Black or African American	1,264	10.4%	21.4%	1.4%	66.9%
Hispanic or Latinx	3,230	15.1%	28.9%	1.4%	54.6%
White	26,263	18.4%	20.2%	0.7%	60.7%
Other	1,462	15.0%	27.4%	1.7%	55.8%
Excluding Missing					
Black or African American	419	31.3%	64.4%	4.3%	-
Hispanic or Latinx	1,466	33.4%	63.6%	3.0%	-
White	10,332	46.8%	51.4%	1.8%	-
Other	646	34.1%	62.1%	3.9%	-

Note. Percentages and total number of teachers apply to each row.

Exhibit 30. 2022–23 Teacher Evaluation Ratings in Schools in the 4th Quartile of Low-Income Enrollment by Teacher Race/Ethnicity

Teacher Race/Ethnicity	Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
Black or African American	5,954	3.4%	18.6%	3.8%	74.1%
Hispanic or Latinx	5,500	7.0%	20.4%	3.0%	69.6%
White	15,899	10.2%	20.4%	1.8%	67.5%
Other	1,696	8.1%	21.3%	3.6%	67.0%
Excluding Missing					
Black or African American	1,540	13.3%	71.9%	14.7%	-
Hispanic or Latinx	1,674	23.0%	67.1%	9.9%	-
White	5,161	31.4%	63.0%	5.6%	-
Other	560	24.5%	64.6%	10.9%	-

Note. Percentages and total number of teachers apply to each row.

Findings from district-characteristic cross tabulations corroborate those from school characteristics: teachers working in contexts with more disadvantage tended to receive lower ratings. Exhibit 31 shows the distribution of evaluation ratings by the districts' quartiles for low-income student enrollment. Among those evaluated in 2022–23, a higher percentage of teachers in districts with higher proportions of low-income students received "Needs Improvement" or "Unsatisfactory" ratings (5.3% for Quartile 4 compared to 1.2% for Quartile 1) and a lower percentage received a rating of "Excellent" (34.1% for Quartile 4 versus 55.5% for Quartile 1). Teacher evaluation rating distributions by the districts' prior-year ELA and math achievement proficiency quartile were also analyzed. Results were similar to the low-income student enrollment cross tabulations: teachers in districts with lower prior-year achievement proficiency rates tended to receive lower ratings.

Exhibit 31. 2022–23 Teacher Evaluation Ratings by District Low-Income Student Enrollment Quartile

District Low-Income Student Enrollment Quartile	Total Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
1st (Percentile 1–25)	35,119	23.0%	18.0%	0.5%	58.5%
2nd (Percentile 26–50)	24,726	17.6%	16.4%	0.5%	65.5%
3rd (Percentile 51–75)	22,397	19.9%	18.7%	0.7%	60.7%
4th (Percentile 76–99)	52,059	11.8%	21.1%	1.8%	65.3%
Multiple or Missing Quartile	4,338	22.9%	18.4%	1.2%	57.4%
Excluding Missing					
1st (Percentile 1–25)	14,581	55.5%	43.4%	1.2%	-
2nd (Percentile 26–50)	8,532	51.0%	47.6%	1.4%	-
3rd (Percentile 51–75)	8,800	50.6%	47.6%	1.8%	-
4th (Percentile 76–99)	18,083	34.1%	60.6%	5.3%	-
Multiple or Missing Quartile	1,847	53.9%	43.2%	2.9%	-

Note. Percentages and total number of teachers apply to each row.

CPS enrolls 17.2% of students and employs 16.4% of teachers in the state. To better understand how ratings of teachers in this unique district compared to ratings of teachers in other districts, we disaggregated data into CPS and non-CPS districts. The ratings of CPS teachers differed from the ratings of teachers in other districts. Only 22.4% of CPS teachers were evaluated in 2022–23 compared to 40.4% of teachers in other districts. Among teachers who were evaluated, CPS teachers also received lower ratings, with only 23.3% rated “Excellent” compared to 48.9% of teachers in other districts, and 13.8% rated “Needs Improvement” or “Unsatisfactory” compared to 1.6% of teachers in other districts.

Exhibit 32. 2022–23 Teacher Evaluation Ratings by Chicago Public Schools

Chicago Public Schools	Total Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
Chicago	22,977	5.2%	14.1%	3.1%	77.6%
Non-Chicago	115,662	19.8%	20.0%	0.7%	59.6%
Excluding Missing					
Chicago	5,141	23.3%	62.9%	13.8%	-
Non-Chicago	46,702	48.9%	49.5%	1.6%	-

Note. Percentages and total number of teachers apply to each row.

AIR disaggregated teacher evaluation ratings by district locale. As shown in Exhibit 33, a higher percentage of city teachers (69.3% vs. 63.8% or lower for other locales) did not receive ratings in 2022–23. Among teachers who were evaluated, a much lower percentage of city teachers were rated as "Excellent" compared with teachers in other locales (30.7% vs. 50.5% or higher), and a higher percentage of city teachers were rated as "Needs Improvement" or "Unsatisfactory" (7.2% vs. 1.8% or lower).

Exhibit 33. Teacher Evaluation Ratings by District Locale

District Locale	Total Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
City	37,462	9.4%	19.1%	2.2%	69.3%
Suburb	71,193	20.8%	19.4%	0.6%	59.2%
Town	15,074	19.0%	16.6%	0.5%	63.8%
Rural	13,407	19.7%	18.6%	0.7%	61.0%
Missing Locale	1,503	12.2%	26.9%	1.1%	59.9%
Excluding Missing					
City	11,493	30.7%	62.2%	7.2%	-
Suburb	29,065	51.0%	47.4%	1.6%	-

District Locale	Total Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Town	5,450	52.7%	46.0%	1.4%	-
Rural	5,232	50.5%	47.7%	1.8%	-
Multiple or Missing Locale	603	30.3%	67.0%	2.7%	-

Note. Percentages and total number of teachers apply to each row.

The distribution of evaluation ratings also differed by the district's teacher characteristics (Exhibit 34 and Exhibit 35). Exhibit 34 shows evaluation ratings by districts' average teacher experience. Among teachers who received a rating, a higher percentage of those in districts with more average years of experience were rated "Excellent" (61.8% for Quartile 4 compared to 37.5% for Quartile 1) and a lower percentage in those districts were rated "Needs Improvement" or "Unsatisfactory."

Exhibit 34. 2022–23 Teacher Evaluation Ratings by District Teacher Years of Experience Quartile

District Teacher Years of Experience Quartile	Total Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
1st (Percentile 1–25)	20,078	15.4%	24.5%	1.2%	58.9%
2nd (Percentile 26–50)	58,798	12.7%	19.5%	1.5%	66.3%
3rd (Percentile 51–75)	29,429	21.7%	18.1%	0.6%	59.6%
4th (Percentile 76–99)	26,676	24.0%	14.5%	0.3%	61.1%
Multiple or Missing Quartile	3,658	19.8%	20.1%	1.4%	58.7%
Excluding Missing					
1st (Percentile 1–25)	8,258	37.5%	59.7%	2.9%	-
2nd (Percentile 26–50)	19,826	37.6%	57.9%	4.6%	-
3rd (Percentile 51–75)	11,879	53.7%	44.8%	1.5%	-
4th (Percentile 76–99)	10,368	61.8%	37.4%	0.8%	-
Multiple or Missing Quartile	1,512	47.9%	48.6%	3.4%	-

Districts were sorted into quartiles by their teacher retention rates, as reported in the ISBE Illinois Report Card data. Among those who were evaluated, a higher percentage of teachers in districts with higher teacher retention rates were rated "Excellent" (63.2% for Quartile 4 versus 35.0% for Quartile 1, Exhibit 35).

Exhibit 35. 2022–23 Teacher Evaluation Ratings by District Teacher Retention Quartile

District Teacher Retention Quartile	Total Number of Teachers	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
1st (Percentile 1–25)	12,934	12.9%	22.9%	1.0%	63.2%
2nd (Percentile 26–50)	49,638	11.9%	20.5%	1.9%	65.7%
3rd (Percentile 51–75)	40,330	20.6%	20.3%	0.6%	58.6%
4th (Percentile 76–99)	32,103	23.3%	13.3%	0.3%	63.1%
Multiple or Missing Quartile	3,634	19.8%	20.0%	1.5%	58.7%
Excluding Missing					
1st (Percentile 1–25)	4,761	35.0%	62.2%	2.8%	-
2nd (Percentile 26–50)	17,041	34.6%	59.8%	5.6%	-
3rd (Percentile 51–75)	16,708	49.6%	49.0%	1.4%	-
4th (Percentile 76–99)	11,832	63.2%	36.1%	0.8%	-
Multiple or Missing Quartile	1,501	48.0%	48.4%	3.5%	-

Note. Percentages and total number of teachers apply to each row.

Principal and Assistant Principal Evaluation Ratings

Findings for the evaluation rating distributions for principal and assistant principals were similar to the findings for teachers. Because the number of principals and assistant principals was much lower (only 7,024 in 2022–23) than the number of teachers, cross tabulations for principal and assistant principal evaluation ratings and some educator, school, and district characteristics resulted in small cell sizes. To ensure data privacy, only cross tabulations for groups with 10 or more principals and assistant principals in every evaluation rating category were reported.

Exhibit 36 below shows the distribution of principal and assistant principal evaluation ratings by gender. Male and female teachers had similar percentages in each rating category.

Exhibit 36. 2022–23 Principal/Assistant Principal Evaluation Ratings by Gender

Gender	Number of Principals and Assistant Principals	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
Male	2,884	14.9%	16.2%	1.2%	67.8%
Female	4,140	15.1%	15.8%	0.9%	68.2%
Excluding Missing					
Male	927	46.2%	50.2%	3.7%	-
Female	1,319	47.5%	49.7%	2.8%	-

Note. Percentages and total number of principals and assistant principals apply to each row.

Rating distributions were different by race/ethnicity (Exhibit 37). In 2022–23, 81.4% of Black or African American principals and assistant principals did not receive a rating compared to 64.1% of White principals and assistant principals. Among those who received a rating, Black or African American and Hispanic or Latinx principals and assistant principals received an "Excellent" rating at lower rates (27.0% and 20.0%, respectively, versus 51.2%) and received "Needs Improvement" or "Unsatisfactory" ratings at higher rates (7.9% and 10.4%, respectively, versus 2.0%) compared to White principals and assistant principals.

Exhibit 37. 2022–23 Principal/Assistant Principal Evaluation Ratings by Race/Ethnicity

Race/Ethnicity	Number of Principals and Assistant Principals	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
Black or African American	1,150	5.0%	12.1%	1.5%	81.4%
Hispanic or Latinx	511	4.9%	16.9%	2.5%	75.7%
White	5,179	18.4%	16.8%	0.7%	64.1%
Other	184	S	S	S	S
Excluding Missing					
Black or African American	215	27.0%	65.1%	7.9%	-

Race/Ethnicity	Number of Principals and Assistant Principals	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Hispanic or Latinx	125	20.0%	69.6%	10.4%	-
White	1,857	51.2%	46.8%	2.0%	-
Other	49	S	S	S	-

Note. Percentages and total number of principals and assistant principals apply to each row. *S* indicates that data were suppressed because at least one cell in the row represents data for fewer than 10 principals and assistant principals.

Evaluation ratings were also disaggregated by years of experience.[10] Exhibit 38 shows the distribution of principal and assistant principal ratings by years of experience in the district. Among those who were evaluated, principals and assistant principals with 6 or more years of experience had much higher rates of receiving "Excellent" ratings (56.5%) compared to first-year principals and assistant principals (27.9%) and second- or third-year principals and assistant principals (38.8%).

Exhibit 38. 2022–23 Principal/Assistant Principal Evaluation Ratings by Years of Experience in District

Years of Experience	Number of Principals and Assistant Principals	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
1st Year	1,264	8.2%	20.1%	1.1%	70.6%
2 or 3 Years	1,353	12.3%	18.0%	1.5%	68.2%
4 or 5 Years	1,151	S	S	S	S
6 or More Years	3,256	18.3%	13.2%	0.9%	67.6%

[10] Bowers and White (2014) found that student test score growth is higher in non-CPS Illinois schools with principals with 2–5 years of experience, compared to first-year principals and principals with 6 or more years of experience. AIR researchers followed Bowers and White (2014) when organizing experience categories and then further differentiated between principals with 2–3 years of experience and 4–5 years of experience.

Years of Experience	Number of Principals and Assistant Principals	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Excluding Missing					
1st Year	373	27.9%	68.4%	3.8%	-
2nd or 3rd Year	430	38.8%	56.5%	4.7%	-
4th or 5th Year	389	S	S	S	-
6th or More Year	1,054	56.5%	40.8%	2.8%	-

Note. Percentages and total number of principals and assistant principals apply to each row. *S* indicates that data were suppressed because at least one cell in the row represents data for fewer than 10 principals and assistant principals.

Cross tabulations of school and district characteristics and principal and assistant principal evaluation ratings offered suggestive evidence for variation by the school's or districts' student demographic, achievement, locale, and teacher characteristics. Findings for school and district characteristics that have three or more groups (rows) with nonsuppressed data for comparison. The findings reported below are similar to the teacher evaluation findings.

Exhibit 39 shows that principals and assistant principals in schools with concentrations of low-income students tended to receive lower ratings compared to principals and assistant principals in districts with lower proportions of low-income students. For example, among those who received a rating, only 22.4% of principals and assistant principals in the highest quartile of low-income student enrollment were rated "Excellent" compared to 54.0% in the second lowest quartile of low-income student enrollment. Results for prior-year (2021–22) school ELA and math achievement proficiency quartiles (not shown) reveal a similar pattern.

Exhibit 39. 2022–23 Principal/Assistant Principal Evaluation Ratings by School Low-Income Student Enrollment Quartile

School Low-Income Student Enrollment Quartile	Total Number of Principals and Assistant Principals	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
1st (Percentile 1–25)	1,721	S	S	S	S
2nd (Percentile 26–50)	1,453	20.5%	16.8%	0.7%	62.1%
3rd (Percentile 51–75)	1,646	13.5%	20.2%	2.1%	64.2%
4th (Percentile 76–99)	1,758	4.1%	13.0%	1.1%	81.8%
Multiple or Missing Quartile	446	S	S	S	S
Excluding Missing					
1st (Percentile 1–25)	648	S	S	S	-
2nd (Percentile 26–50)	550	54.0%	44.2%	1.8%	-
3rd (Percentile 51–75)	590	37.8%	56.4%	5.8%	-
4th (Percentile 76–99)	322	22.4%	71.4%	6.2%	-
Multiple or Missing Quartile	136	S	S	S	-

Note. Percentages and total number of principals and assistant principals apply to each row. *S* indicates that data were suppressed because at least one cell in the row represents data for fewer than 10 principals and assistant principals.

Consistent with findings for school characteristics, the distribution of evaluation ratings for principals and assistant principals also differed by district characteristics. Principals and assistant principals working in districts with higher average teacher experience and higher district teacher retention rates tended to have higher rates of receiving an “Excellent” rating. For teacher experience, 52.7% of evaluated principals and assistant principals in Quartile 3 districts received ratings of “Excellent” compared to 31.1% for Quartile 1 districts (Exhibit 40).

Exhibit 40. 2022–23 Principal/Assistant Principal Evaluation Ratings by District Teacher Years of Experience Quartile

District Teacher Years of Experience Quartile	Total Number of Principals and Assistant Principals	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
1st (Percentile 1–25)	1,104	9.3%	19.4%	1.3%	70.0%
2nd (Percentile 26–50)	3,113	10.2%	14.5%	1.4%	73.8%
3rd (Percentile 51–75)	1,427	20.8%	18.0%	0.7%	60.4%
4th (Percentile 76–99)	1,100	S	S	S	S
Multiple or Missing Quartile	280	S	S	S	S
Excluding Missing					
1st (Percentile 1–25)	331	31.1%	64.7%	4.2%	-
2nd (Percentile 26–50)	816	39.1%	55.5%	5.4%	-
3rd (Percentile 51–75)	564	52.7%	45.6%	1.8%	-
4th (Percentile 76–99)	450	S	S	S	-
Multiple or Missing Quartile	85	S	S	S	-

Note. Percentages and total number of principals and assistant principals apply to each row. *S* indicates that data were suppressed because at least one cell in the row represents data for fewer than 10 principals and assistant principals.

For teacher retention, 50.1% of evaluated principals and assistant principals in Quartile 3 districts received a rating of “Excellent” compared to 30.9% for Quartile 1 (Exhibit 41).

Exhibit 41. 2022–23 Principal/Assistant Principal Evaluation Ratings by District Teacher Retention Quartile

District Teacher Retention Quartile	Total Number of Principals and Assistant Principals	Excellent	Proficient	Needs Improvement or Unsatisfactory	Missing Rating
Including Missing					
1st (Percentile 1–25)	799	7.3%	14.8%	1.5%	76.4%
2nd (Percentile 26–50)	2,687	7.0%	14.0%	0.9%	78.1%
3rd (Percentile 51–75)	1,956	21.2%	19.5%	1.6%	57.6%
4th (Percentile 76–99)	1,305	S	S	S	S
Multiple or Missing Quartile	277	S	S	S	S
Excluding Missing					
1st (Percentile 1–25)	188	30.9%	62.8%	6.4%	-
2nd (Percentile 26–50)	589	32.1%	64.0%	3.9%	-
3rd (Percentile 51–75)	828	50.1%	46.1%	3.7%	-
4th (Percentile 76–99)	556	S	S	S	-
Multiple or Missing Quartile	85	S	S	S	-

Note. Percentages and total number of principals and assistant principals apply to each row. S indicates that data were suppressed because at least one cell in the row represents data for fewer than 10 principals and assistant principals.

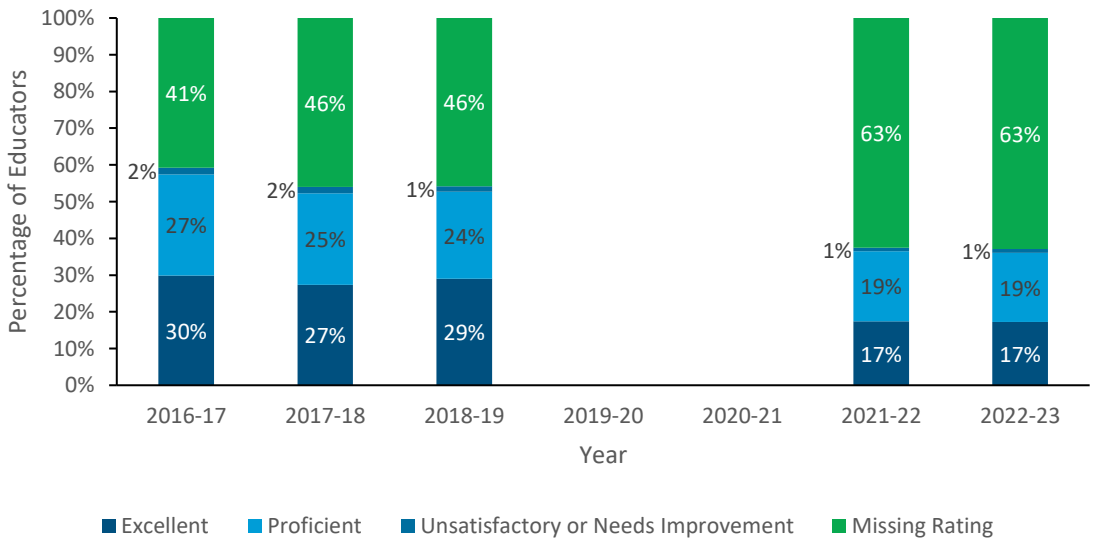
RQ6c. How did the distributions of evaluation ratings change over time?

ISBE educator-level data for school years 2013–14 to 2022–23 were used to answer RQ6c. Because data for 2013–14 to 2015–16 had missing values for evaluation ratings for 80% or more of the educators in each year, we used data for 2016–17 to 2022–23 for the analysis. Also excluded from the analysis were the 2019–20 and 2020–21 school years, because educator evaluation in those 2 years was affected by the COVID-19 pandemic. Each year of data included records for 139,280 to 145,740 educators. Teachers, principals, and assistant principals were combined for the analyses.

For teachers, principals, and assistant principals with one evaluation rating record in each year, the percentage of educators who received "Excellent," "Proficient," "Needs Improvement," or "Unsatisfactory" ratings or who were coded "Missing Rating" was calculated. The distribution of ratings was examined for all teachers, principals, and assistant principals in each year, as well as by educator race/ethnicity, school level, school low-income student enrollment quartile, and whether the school was in CPS. Below, the combined percentages of teachers, principals, and assistant principals who received "Needs Improvement" and "Unsatisfactory" ratings were reported.

Exhibit 42 shows the distribution of evaluation ratings from 2016–17 to 2022–23. While the percentage of teachers, principals, and assistant principals who were evaluated decreased over time (59.3% in 2016–17 versus 37.1% in 2022–23), the distribution of ratings among those who were evaluated were similar across years. The percentage of educators who received an "Excellent" or "Proficient" rating ranged between 96.8% and 97.5% during this time, while between 2.5% and 3.2% of educators received "Needs Improvement" or "Unsatisfactory" ratings from year to year.

Exhibit 42. Evaluation Rating Distributions 2016–17 to 2022–23, All Teachers, Principals, and Assistant Principals



The 2016–17 to 2022–2023 evaluation rating data was disaggregated by educator race/ethnicity, school level, school quartile for the proportion of low-income students enrolled, and whether the school was in the CPS system. Across the whole state, a smaller percentage of teachers, principals, and assistant principals were evaluated in 2022–23 compared to 2016–17. Differences over time for the evaluation distributions differed by educator and school characteristics. Among Black or African American teachers, principals, and assistant principals

who received ratings, the percentage rated “Excellent” or “Proficient” increased by 3.5 percentage points. Among Hispanic or Latinx teachers, principals, and assistant principals who received ratings, the percentage rated “Excellent” decreased by 9.5 percentage points while the percentage rated “Proficient” increased by 9.1 percentage points. The distribution for White educators stayed the same. Elementary and middle school educators saw a decrease in the percentage of “Excellent” ratings and a corresponding increase in the percentage of “Proficient” ratings, while the distribution for high school educators stayed the same. Evaluation rating distributions for teachers in schools with the lowest percentage of low-income students did not change much over time, whereas the percentage of “Excellent” ratings decreased, and the percentage of “Proficient” ratings increased considerably for teachers in schools with the highest percentage of low-income students. For teachers in both CPS and non-CPS schools, the percentage of “Excellent” ratings decreased while the percentage of “Proficient” ratings increased, but the change was more substantial for CPS teachers. The results are shown in Exhibit 43 through Exhibit 45.

Exhibit 43 shows the distribution of evaluation ratings for Black or African American teachers, principals, and assistant principals. As for all educators, the percentage of missing ratings increased between 2016–17 (42.2%) and 2022–23 (72.2%). Among all Black or African American educators, the percentage rated “Excellent” or “Proficient” decreased from 49.7% to 24.9%. Among Black or African American educators who received ratings, the percentage rated “Excellent” or “Proficient” increased from 86.1% in 2016–17 to 89.6% in 2022–23.

Exhibit 43. Evaluation Rating Distributions 2016–17 to 2022–23, Black or African American Educators

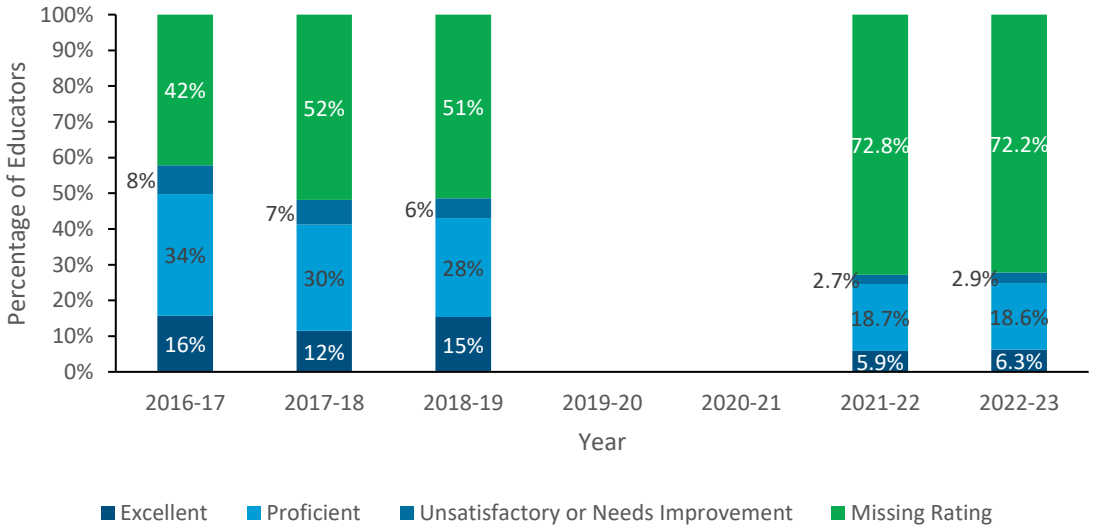
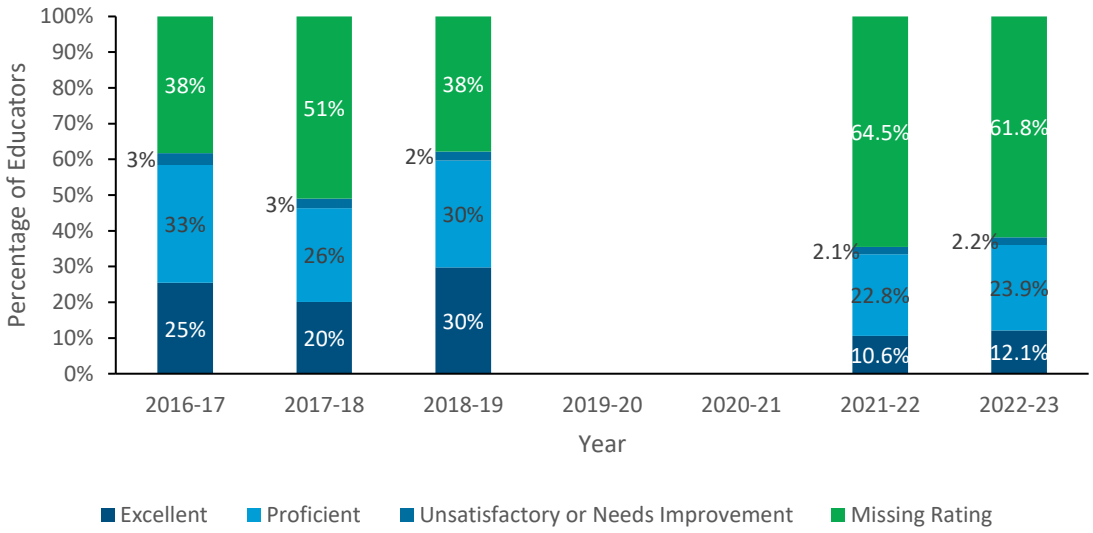


Exhibit 44 shows the distribution of evaluation ratings for Hispanic or Latinx teachers, principals, and assistant principals. As for all educators, the percentage of missing ratings increased between 2016–17 (38.2%) and 2022–23 (61.8%). Among all Hispanic or Latinx educators, the percentage rated "Excellent" or "Proficient" decreased from 58.4% to 36.0%; the percentage rated "Needs Improvement" or "Unsatisfactory" decreased from 3.3% to 2.2%. Among Hispanic or Latinx educators who received ratings, the percentage rated "Excellent" decreased by 9.5 percentage points (41.2% to 31.7%) while the percentage rated "Proficient" increased by 9.1 percentage points (53.5% to 62.6%). The percentage of Hispanic or Latinx educators rated "Needs Improvement" or "Unsatisfactory" was similar (5.4% versus 5.6%) between 2016–17 and 2022–23.

Exhibit 44. Evaluation Rating Distributions 2016–17 to 2022–23, Hispanic or Latinx Educators



As shown in Exhibit 45, White educators' distribution of evaluation ratings was stable from 2016–17 to 2022–23. The percentage of White teachers, principals, and assistant principals missing ratings went up from 40.8% to 61.6%. Among all White educators, the percentage rated "Excellent" or "Proficient" decreased from 57.9% to 36.8%; the percentage rated "Needs Improvement" or "Unsatisfactory" decreased from 1.2% to 0.7%. Among White educators who received ratings, distribution across rating categories stayed approximately the same. The percentage of White educators rated "Excellent" decreased from 53.3% to 50.1% and the percentage of "Proficient" ratings increased from 44.7% and 48.0%; the combination of "Needs Improvement" and "Unsatisfactory" ratings constituted 2.1% in 2016–17 and 2.0% in 2022–23.

Exhibit 45. Evaluation Rating Distributions 2016–17 to 2022–23, White Educators

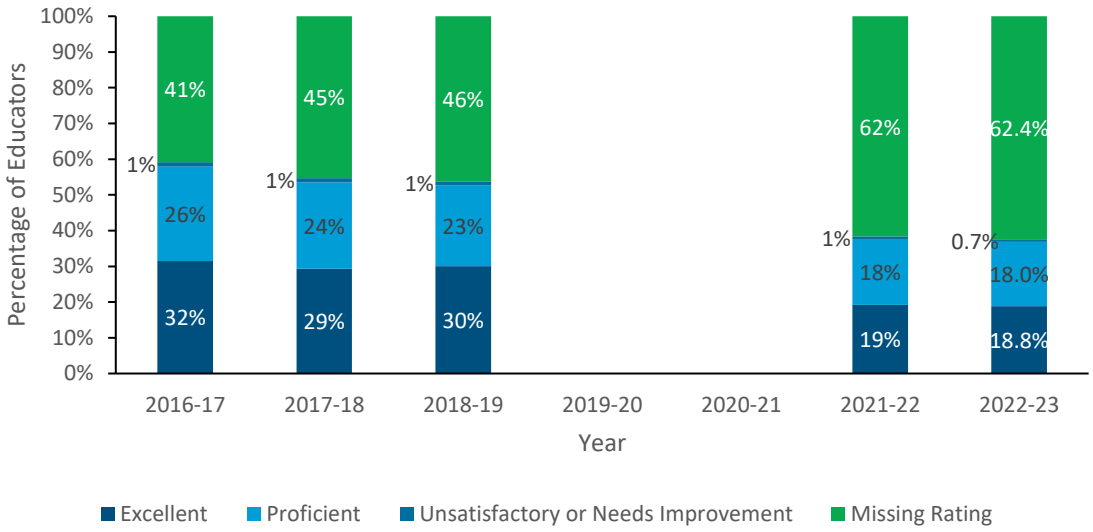


Exhibit 46 through Exhibit 48 show the distribution of educator evaluation ratings over time by school level. The percentage of educators who did not receive a rating increased for all three school levels: 38.0% to 63.4% for elementary schools (Exhibit 46); 40.7% to 59.8% for middle schools (Exhibit 47); and 42.9% to 64.2% for high schools (Exhibit 48). The percentage of educators who received each category of rating decreased corresponding to this increase in missing data. Among elementary school educators who received ratings, the percentage rated "Unsatisfactory" or "Needs Improvement" largely stayed the same from 2016–17 (3.8%) to 2022–23 (3.5%), whereas the percentage of educators rated "Proficient" increased (46.3% to 52.7%) and the percentage rated "Excellent" decreased (49.9% to 43.7%). A similar trend is seen for middle school educators who received ratings. Among high school educators who received ratings, the percentage for each rating category stayed approximately the same over time.

Exhibit 46. Evaluation Rating Distributions 2016–17 to 2022–23, Educators in Elementary Schools

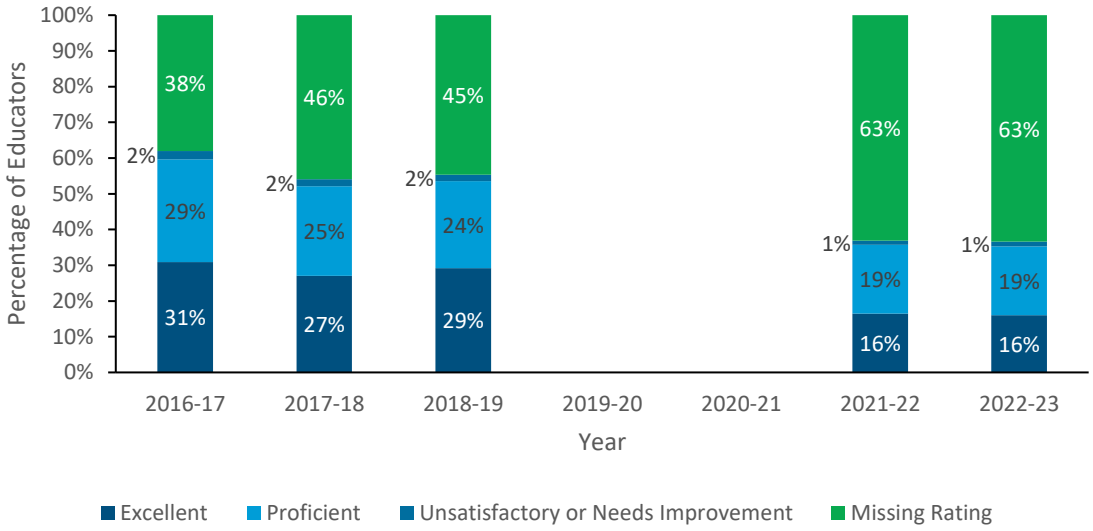


Exhibit 47. Evaluation Rating Distributions 2016–17 to 2022–23, Educators in Middle Schools

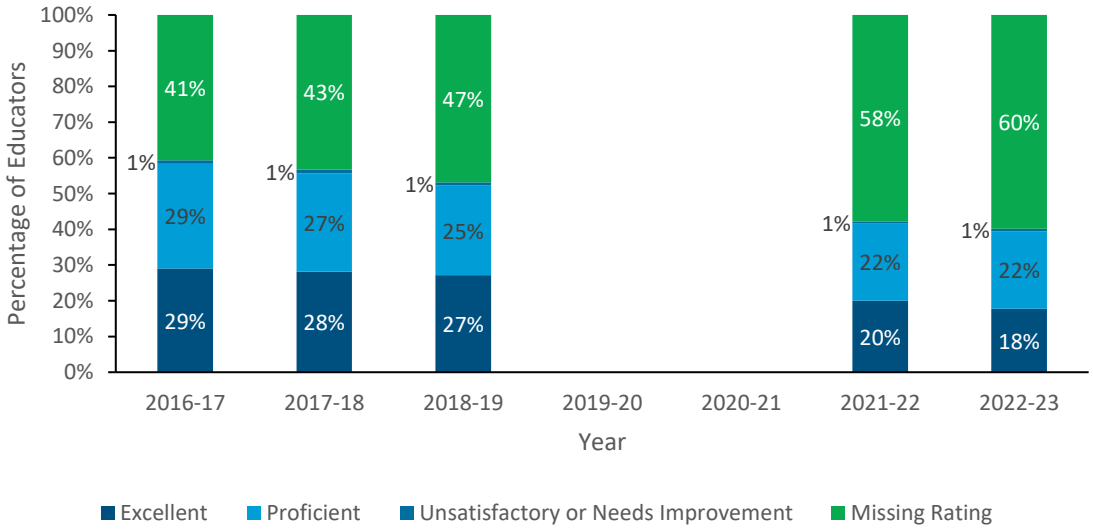
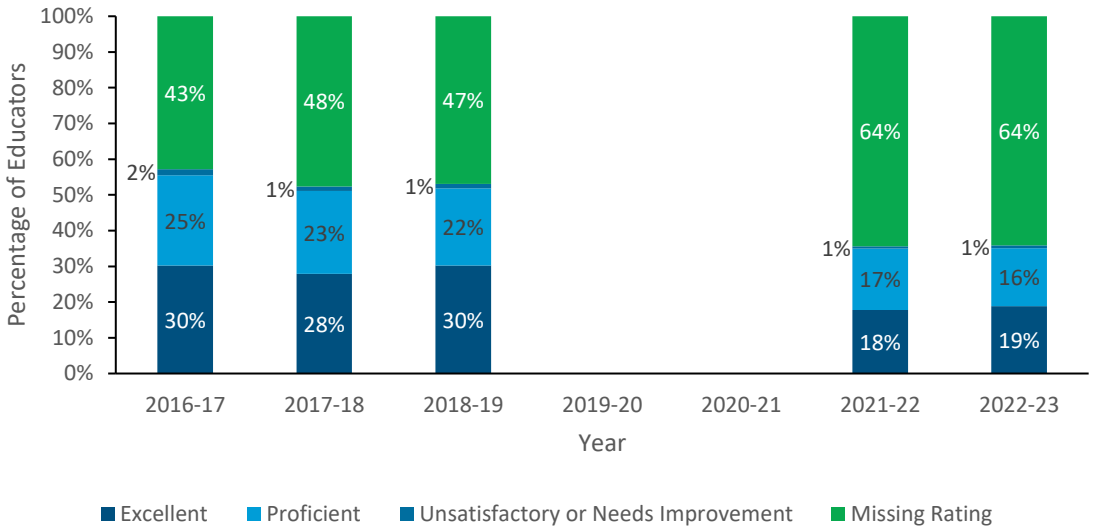


Exhibit 48. Evaluation Rating Distributions 2016–17 to 2022–23, Educators in High Schools



The changes over time for the distribution of evaluation ratings differed more by school quartile of low-income student enrollment. Quartiles 1 and 4 are shown below. Across all four quartiles, the percentage of educators missing ratings increased (42.4% to 60.7% for Quartile 1; 36.7% to 69.9% for Quartile 4); the percentage of educators receiving each category of ratings decreased corresponding to this increase in missing data.

For schools with the lowest percentage of low-income students (Quartile 1, Exhibit 49), the percentage of educators rated “Excellent” decreased by 3.3 percentage points and the percentage rated “Proficient” increased by 3.1 percentage points. In contrast, educators who received ratings in schools with the highest percentages of low-income students (Quartile 4,Exhibit 50) saw much larger decreases in the percentage of educators receiving “Excellent” ratings (34.4% to 26.2%) and larger increases in the percentage receiving “Proficient” ratings (56.4% to 65.6%).

Exhibit 49. Evaluation Rating Distributions 2016–17 to 2022–23, Educators in Low-Income Student Enrollment Quartile 1 Schools

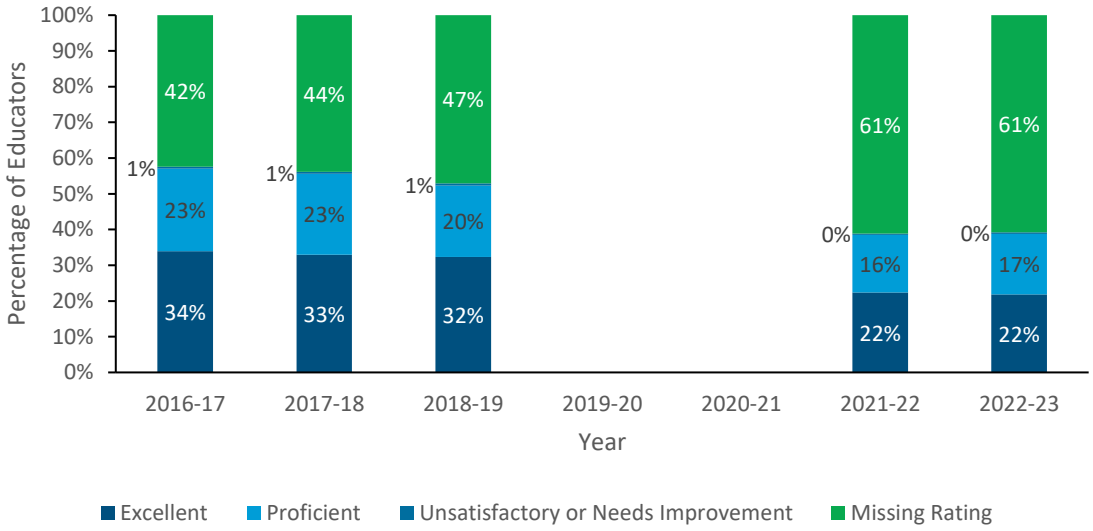


Exhibit 50. Evaluation Rating Distributions 2016–17 to 2022–23, Educators in Low-Income Student Enrollment Quartile 4 Schools

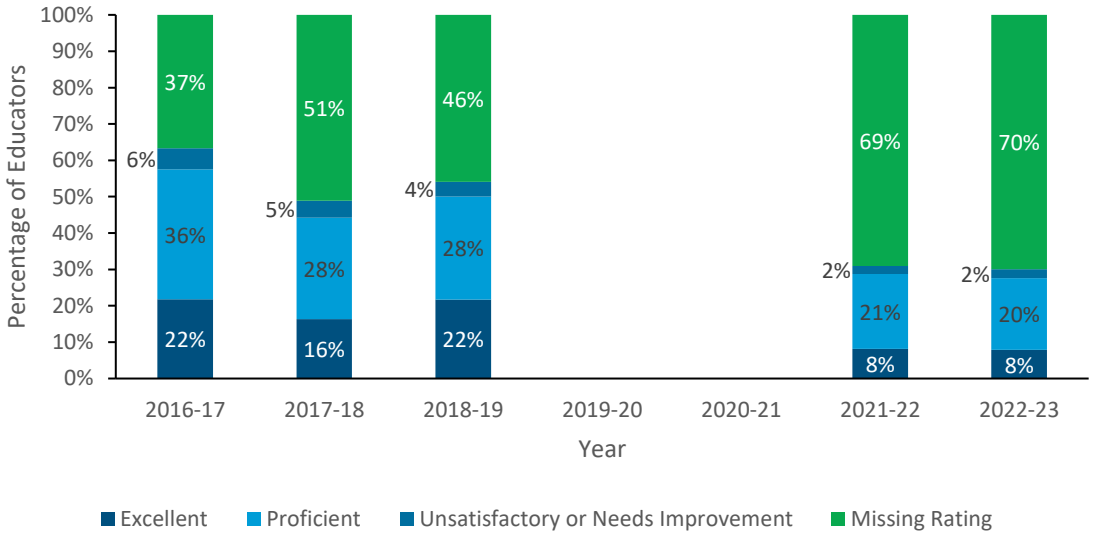


Exhibit 51 and Exhibit 52 show the distribution of evaluation ratings over time for CPS and non-CPS schools. From 2016–17 to 2022–23, the percentage of educators receiving a rating decreased for both CPS (62.6% to 21.2%) and non-CPS schools (58.6% to 40.3%). The percentage of educators who received each category of ratings decreased accordingly. Of the educators who received an evaluation rating, the percentage of "Excellent" ratings decreased while the percentage of "Proficient" ratings increased for both CPS and non-CPS schools, but

the changes were greater for CPS schools (40.3% to 23.3% for "Excellent"; 48.4% to 62.9% for "Proficient").

Exhibit 51. Evaluation Rating Distributions 2016–17 to 2021–22, Educators in Chicago Public Schools

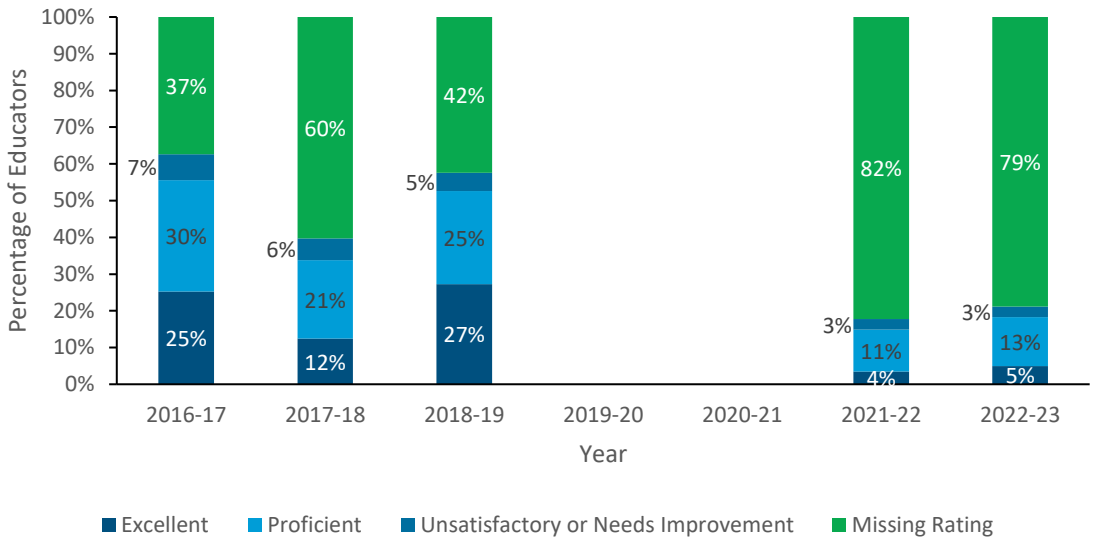
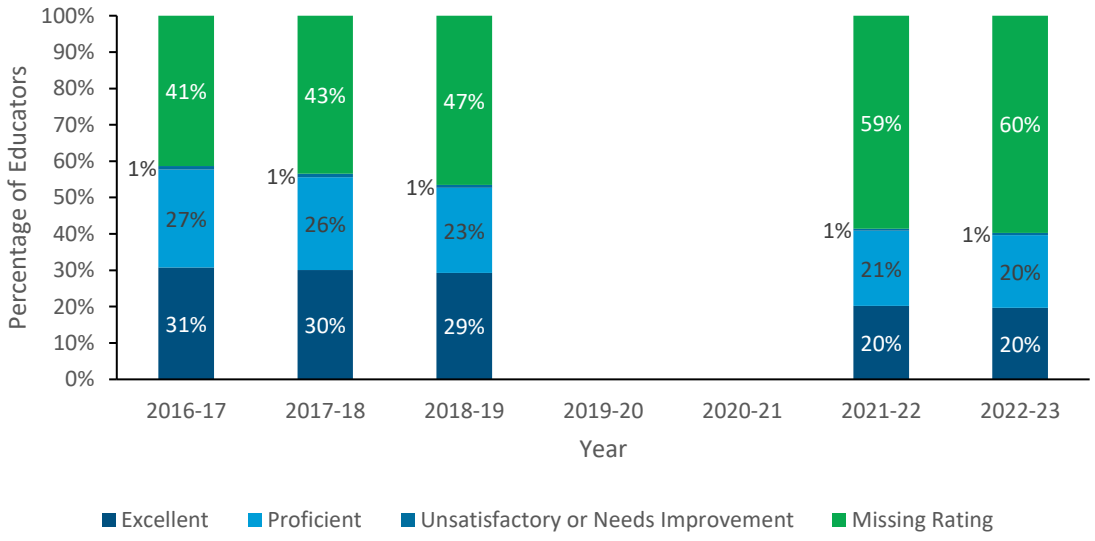


Exhibit 52. Evaluation Rating Distributions 2016–17 to 2022–23, Educators Not in Chicago Public Schools



Summary

To answer RQ6, "What are the distributions of educator ratings across districts and schools and over time, and to what extent do educator ratings vary systematically with respect to the

characteristics of educators, districts, and schools?" the following set of secondary RQs was developed:

- RQ6a. What percentage of educators received evaluation ratings in each of the four categories ("Excellent," "Proficient," "Needs Improvement," and "Unsatisfactory")?
- RQ6b. What percentage of educators received ratings in each of the categories by teacher, school, and district characteristics?
- RQ6c. How did the distributions of evaluation ratings change over time?

Data used in response to the aforementioned included ISBE educator-level data and reported the distribution of educator ratings across districts and schools for 2016–17 to 2022–23.

Of the educators who received an evaluation rating in 2022–23, more than 97% of teachers and more than 96% of principals and assistant principals received a rating of "Excellent" or "Proficient."

The distribution of evaluation ratings varied by educator race/ethnicity and years of experience, as well as by school and district characteristics. Lower percentages of educators of color, educators with fewer years of experience, and educators in schools and districts with more disadvantages were rated "Excellent" or "Proficient."

Survey Analysis

As one of multiple data sources designed to capture perceptions of educators who participated in their district's educator evaluation system during the 2022–23 school year, a survey developed in concert with PEAC and ISBE was administrated to gather insights and experiences from PK–12 public school administrators and teachers (the "population") regarding evaluation systems in Illinois districts. The survey asked questions about different aspects of the evaluation process, such as educators' understanding of the district's evaluation system, coaching opportunities, written feedback, formal observation component, student assessment component, and supports.[11] A complete copy of the survey instrument is presented in Appendix A. Findings from the survey, coupled with interview and focus group data, can provide guidance to state leaders on developing and implementing improved educator evaluation systems and supports.

In the following sections, the methods and procedures for this educator survey administration and data analysis are described, and the main findings are shared.

Sample Selection

The population comprises more than 100,000 educators across about 3,800 schools in about 850 districts. A multistage, stratified, systematic sampling approach was employed to ensure a representative sample from the full population. First, 125 districts were selected at random from strata—groups of districts with similar characteristics—defined by locale, ROE service area,[12] and the percentage of students eligible for the federal free or reduced-price lunch (FRPL) program in each district. This stratification is to ensure that different subgroups (strata) within the population are adequately represented in the sample based on these characteristics. Within each stratum, districts were sorted by enrollment size and racial demographics before the systematic sample was taken. Systematic sampling involves choosing a random starting point from the sorted list and then selecting every *m*th record thereafter. The interval between selections, *m*, was calculated by dividing the total population size by the desired sample size within each stratum. For school selection, 300 schools were chosen at random from the sample districts. Within each randomly selected district, schools were sorted by Title I status, locale, and student demographic characteristics before systematic selection within each district. All teachers and administrators within the selected schools were invited to participate in the survey through emails sent directly to them. The sampling approach ensures representation

[11] The survey used "student assessment component" as a synonym for the "student growth component" and "formal observation component" as a synonym for "professional practice component." Because survey respondents might not have considered these to be synonyms, we report findings based on the language in the survey.
[12] See Illinois Association of Regional School Superintendents (2015).

across various district and school characteristics, allowing for valid inferences about the broader population while minimizing the burden on educators and logistical challenges of data collection. A more detailed description of the sample selection is presented in Appendix B.

Survey Programming and Administration

A copy of the survey instrument is presented in Appendix A. The survey was administered using an online platform, SurveyMonkey, to ensure ease of use and accessibility, and the survey was accessible on a range of devices, including smartphones or tablets. Potential respondents were contacted about the opportunity to share their perceptions on and experience with educator evaluation using the platforms SurveyMonkey and Airtable, and the research team regularly monitored response rates and followed up with nonrespondents.

The survey included skip logic to capture relevant data for different groups of respondents or types of responses. The skip logic ensured that participants only answered questions applicable to them based on their previous answers. Prior to the survey launch, the survey was tested on various devices and browsers to verify functionality and ensure proper question display and accurate capture of responses. Feedback from the testing phase informed necessary adjustments to aid in ease of access and completion of the instrument.

Data Collection

On November 3, 2023, an email was sent to each school-based administrator and teacher in schools that had been randomly selected to participate in the survey, informing these educators that they would soon receive an email invitation with a link to the survey. On November 7, 2023, educators received a personalized email invitation with a unique link to the survey. These personalized email invitations provided educators with a clear and concise explanation of the survey's purpose, emphasized the importance of their feedback, and assured the anonymity and confidentiality of their responses.

Email notifications and invitations were sent to 11,700 educators located in 298 schools across 124 districts.[13] The survey window commenced November 7, 2023, and concluded December 15, 2023, to provide educators with 6 weeks to respond. During this time frame, personalized follow-up emails to nonrespondents were sent on eight occasions.

The final item on the survey asked teachers and school-based administrators if they would be interested in volunteering for a focus group. From among those who answered "Yes" to the last

[13] The sampling frame was based on schools in operation in 2022–23, but two of these schools had closed by 2023–24, the year in which the survey was administered.

item, a sample was selected to participate in the focus group study, which is reported in the section Interview and Focus Group Analysis.

Nonresponse Bias Analysis and Weighting Adjustments

The survey was administered to a representative sample of educators in Illinois, but not all educator types were equally likely to be eligible or likely to participate. For example, educators with less than 3 years of experience participated at higher rates than their peers with 4 or more years of experience. This may be expected, given that educators with less than 3 years of experience were more likely to receive evaluations than their peers with more experience in 2022–23. Similarly, educators in rural schools responded at a higher rate than educators in city schools.

To account for these differential eligibility and nonresponse patterns among different subgroups, sampling weights (base weights), which were calculated as the inverse of the probability that each educator was selected into the sample, were adjusted. To adjust the sampling weights, logistic regression models were constructed using school and educator characteristics as predictors to estimate the eligibility and response propensities for each sampled individual. Sampling weights were adjusted based on the predicted propensities. These adjusted weights corrected for the differential eligibility and nonresponse patterns among educators in different subgroups and with different demographic characteristics.

The analysis of respondent data then weighted each response so that the responses from the sample would provide an unbiased estimate for the full population. The full population includes all teachers, principals, and assistant principals in Illinois who received evaluations in 2022–23. A more detailed description of the nonresponse bias analysis and weighting adjustments is presented in Appendix E: Survey Weighting Adjustments.

Only educators who were evaluated in an Illinois public school during the school year of 2022–23 and still in the Illinois public school system at the time of the survey were eligible to participate. However, educators who were not formally evaluated in 2022–23 may still have chosen to participate in the survey.

Data Cleaning and Analysis

Prior to analyzing the survey data, these steps were followed to clean and validate the data. First, submitted responses were checked and fixed to ensure they accurately followed the intended skip logic. This “fixing” process addressed instances where respondents went back to earlier sections of the survey to change their answers, sometimes disrupting the intended flow of questions and leading to inconsistencies in the data. Next, analyses were limited to include only respondents who met two criteria: (a) they indicated they were evaluated in an Illinois

public school as a tenured or nontenured teacher, principal, assistant principal, or other school-based administrator during the school year (2022–23), and (b) they provided a valid response (not missing or "Unsure") to any of the questions about educator evaluation systems in their district.

The analysis primarily employed descriptive statistics—tabulations and cross tabulations—to answer the RQs. Statistical testing was employed to ascertain whether the disparities observed in the data were statistically significant at a 95% confidence level. The statistical software Stata's built-in "svyset" and "svy: tabulate" commands (StataCorp, 2023) accounted for key elements of the sample design, such as stratification, clustering, and weighting, in descriptive statistics estimation, statistical testing, and variance estimation through linearization.

To protect personally identifiable information, data for any demographic group with fewer than 10 individuals was suppressed (U.S. Department of Education, 2023, p. 66). This follows National Center for Education Statistics (NCES) best practices for protecting personally identifiable information (Seastrom, 2010, p. 14). In accordance with these guidelines, we suppressed key statistics for subgroups smaller than 10 individuals or larger than $N - 10$ individuals, where N represents the total number of individuals in the sample. This approach ensures that statistics from neither very large nor very small subgroups can be analyzed in ways that might compromise confidentiality. For instance, in a sample of 100 individuals, statistics would be suppressed for any subgroup with fewer than 10 individuals and for any subgroup with more than 90 individuals, thereby preventing indirect disclosure about the remaining 10 individuals.

Characteristics of Survey Respondents

In this section, the characteristics of the 2,588 educators who responded to and were eligible for the survey and the characteristics of educators who are in the target population, including all Illinois educators eligible for the survey are summarized. The sample distributions closely reflect the eligible population, as shown below, which provides confidence that the findings from the survey analysis represent the diverse demographic and professional characteristics of the population. Below, survey respondents' race/ethnicity, gender, work role in 2022–23, years of experience in that work role, school locale, and school level are reported. Percentages reported in the tables below are weighted to reflect the characteristics of the population of Illinois teachers and school leaders.

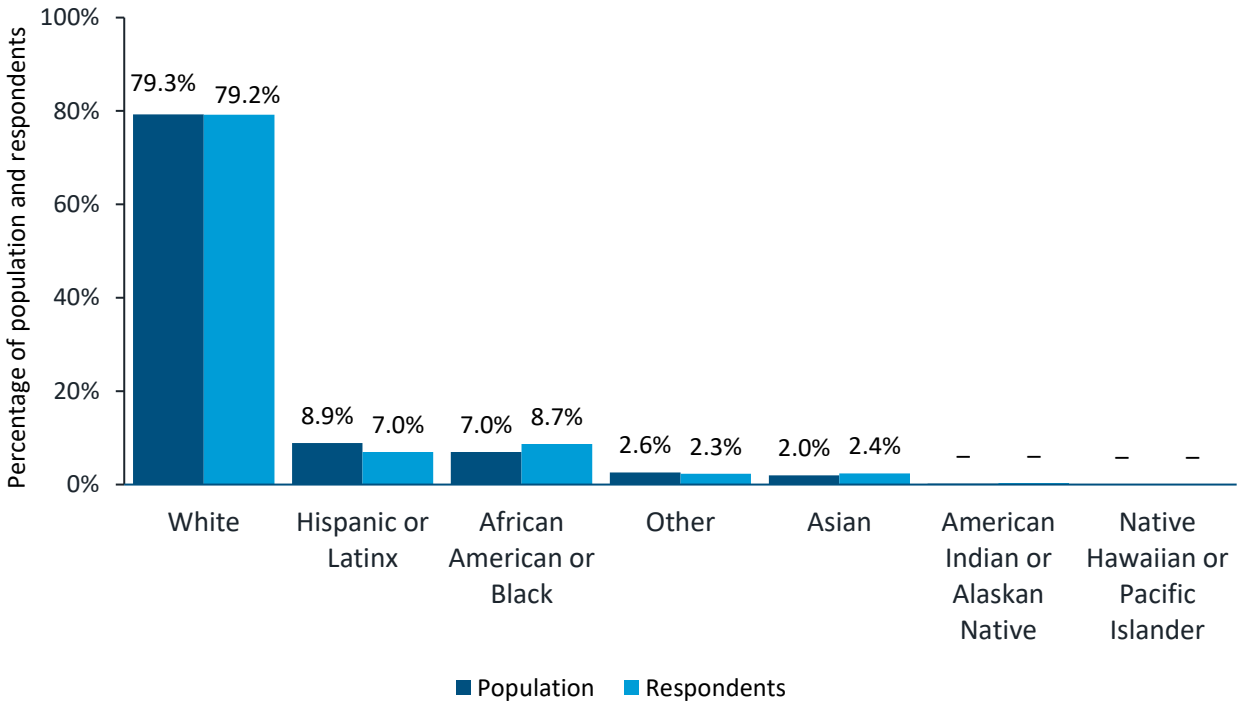
Exhibit 53 shows the racial and ethnic composition of all educators in Illinois and of survey respondents. In terms of race/ethnicity, the majority of survey respondents identified as White (79.2%), followed by Black or African American (8.7%), Hispanic or Latinx (7.0%), Asian (2.4%), and other categories (2.3%). The distribution of respondents across educator categories in our

sample closely mirrored the demographics of all Illinois educators in the population, with all group differences falling within a two-percentage-point margin. Percentages for American Indian, Alaskan Native, Native Hawaiian, and Pacific Islander respondents are not reported because there were fewer than 10 respondents from each of those groups.

Exhibit 53. Race/Ethnicity of Population and Survey Respondents

Race/Ethnicity	Number in Population	Percent in Population	Number of Respondents	Percent of Respondents
Black or African American	6,858	7.0%	124	8.7%
American Indian or Alaskan Native	198	0.2%	–	–
Asian	1,974	2.0%	62	2.4%
Hispanic or Latinx	8,751	8.9%	125	7.0%
Native Hawaiian or Pacific Islander	66	0.1%	–	–
White	78,112	79.3%	2,213	79.2%
Other	2,598	2.6%	54	2.3%
Total	98,557	100.0%	2,588	100.0%

Note. Population data from Illinois State Board of Education (2023). Percentages are weighted. Details may not add up to total due to rounding. Statistics are not shown (–) because the group being analyzed is too small.



Note. Data not shown (–) because the group being analyzed is too small.

Exhibit 54 provides an overview of the gender distribution among all eligible Illinois educators and among survey respondents. Gender statistics are reported in two ways. The gender of the respondents according to ISBE data is reported, and then the gender that respondents indicated in the survey ("Survey Data") is reported. More than three quarters of respondents were female; less than one fourth of the respondents were male; and less than 1.0% of respondents were transgender, nonbinary, or another gender.

Exhibit 54. Gender of Population and Survey Respondents

Gender	Number in Population	Percent in Population	Number of Respondents	Percent of Respondents	Number of Respondents (Survey Data)	Percent of Respondents (Survey Data)
Male	24,459	24.8%	667	22.0%	592	22.6%
Female	74,098	75.2%	1,921	78.0%	1,675	77.6%
Transgender, nonbinary, or another gender	NA	NA	NA	NA	12	0.6%
Total	98,557	100.0%	2,588	100.0%	2,279	NA

Note. Population data from Illinois State Board of Education (2023). NA = Not available. Survey respondents who did not select a gender were excluded. The survey question was "Select all that apply." Percentages are weighted. Details may not add up to total due to rounding.

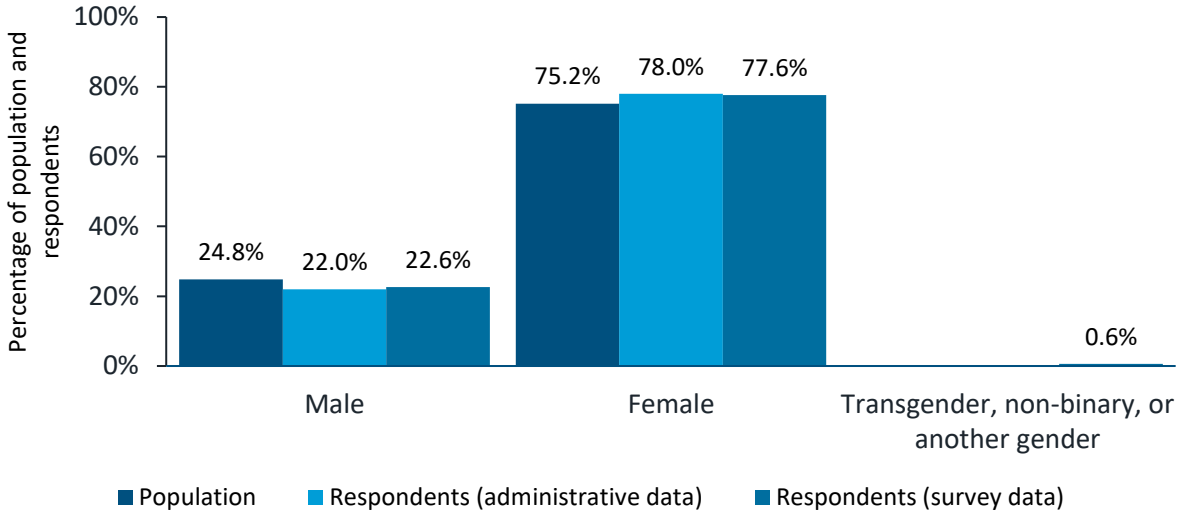


Exhibit 55 presents information on the professional roles of Illinois educators and survey respondents in 2022–23. Professional role statistics are reported in two ways. The role of the respondents according to ISBE administrative data is presented first, and then the role respondents indicated in the survey ("Survey Data"), which included a distinction between tenured and nontenured teachers, is shared. According to ISBE administrative data, the majority of respondents were teachers (96.0%), with a smaller percentage being principals or assistant principals (4.0%).

Exhibit 55. Role of Population and Survey Respondents

Role in 2022–23	Number in Population	Percent in Population	Number of Respondents	Percent of Respondents	Number of Respondents (Survey Data)	Percent of Respondents (Survey Data)
Nontenured teacher	NA	NA	NA	NA	842	40.5%
Tenured teacher	NA	NA	NA	NA	1,303	53.2%
Teacher	95,099	96.5%	2,473	96.0%	NA	NA
Principal/ Assistant principal	3,458	3.5%	115	4.0%	103	4.0%
Other school-based administrator	NA	NA	NA	NA	38	1.1%
Related service provider (e.g., school counselor, nurse)	NA	NA	NA	NA	23	1.2%
Total	98,557	100.0%	2,588	100.0%	2,309	100.0%

Note. Population data from Illinois State Board of Education (2023). NA = Not available. Survey respondents who did not select a role were excluded. Percentages are weighted. Details may not add up to total due to rounding.

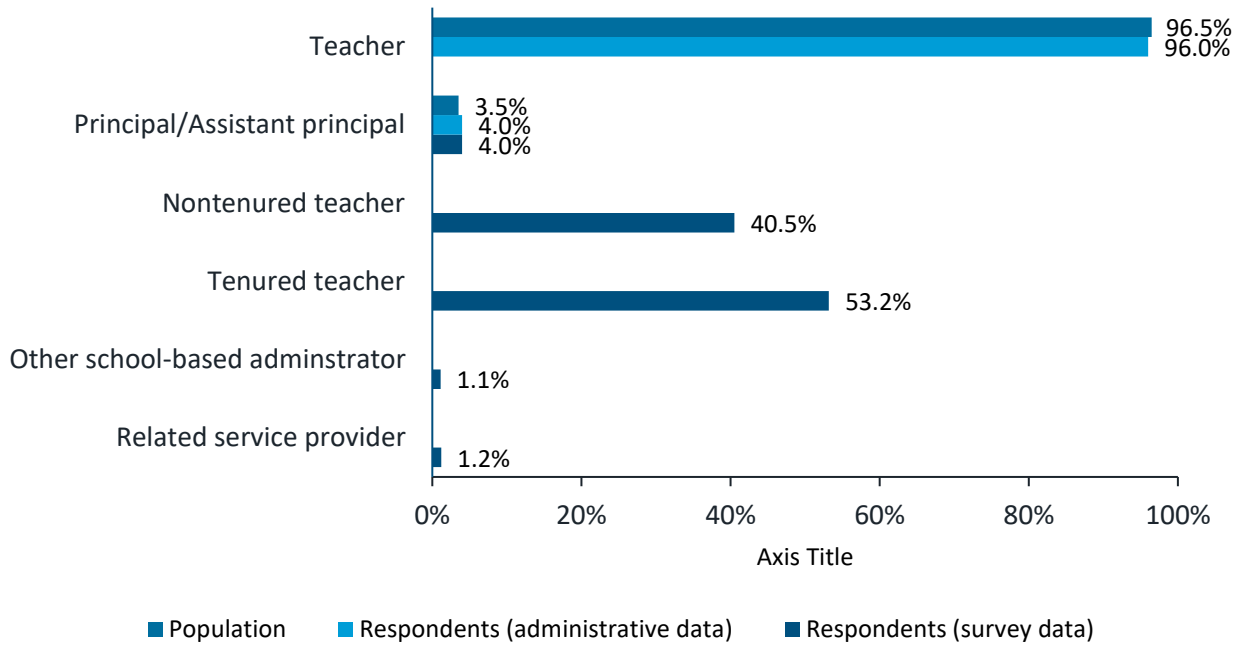


Exhibit 56 presents the distribution of the years of experience of the Illinois educator population and survey respondents in their work roles at the end of 2022–23. Survey respondents were distributed across a wide range of years of experience in their roles. Approximately 21.3% of respondents were in their first year, 20.6% in their second or third year, and 12.9% in their fourth or fifth year; 45.2% had 6 or more years of experience.

Exhibit 56. Years of Experience in Current Role: Population and Survey Respondents

Years of Experience in Role at End of 2022–23	Number in Population	Percent in Population	Number of Respondents	Percent of Respondents
1 year	17,425	17.7%	428	21.3%
2–3 years	20,557	20.9%	501	20.6%
4–5 years	13,429	13.6%	307	12.9%
6 or more years	47,146	47.8%	1,352	45.2%
Total	98,557	100.0%	2,588	100.0%

Note. Population data from Illinois State Board of Education (2023). Survey respondents who did not select years of experience were excluded. Percentages are weighted. Details may not add up to total due to rounding.

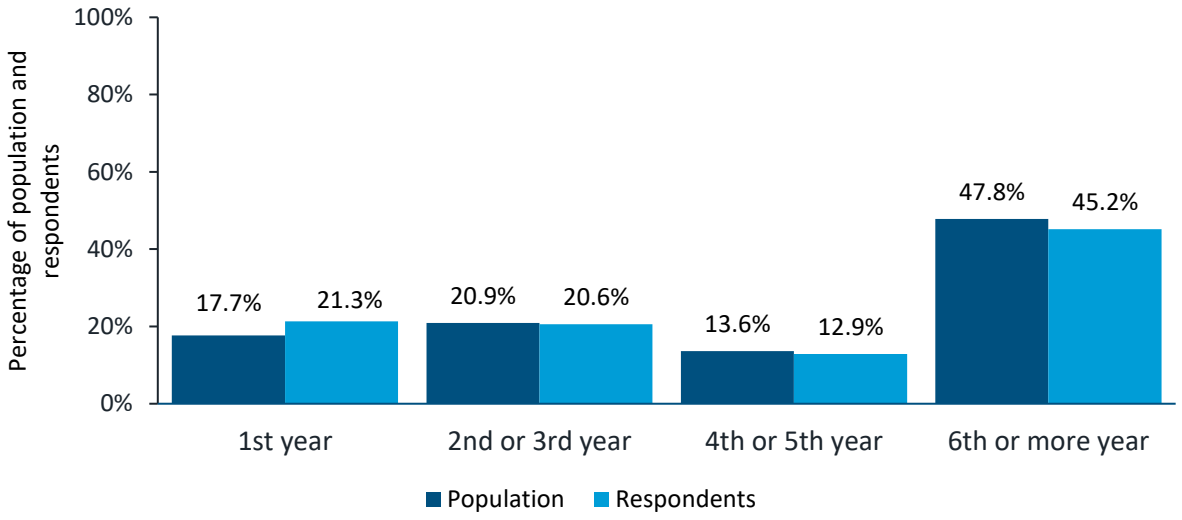


Exhibit 57 presents the locale of the school where the population of Illinois educators and survey respondents worked in 2022–23. The majority of survey respondents worked in schools located in suburbs (54.9%), followed by cities (22.4%), rural areas (12.5%), and towns (10.1%).

Exhibit 57. School Locale: Population and Survey Respondents

School Locale	Number in Population	Percent in Population	Number of Respondents	Percent of Respondents
City	28,952	29.4%	584	22.4%
Suburb	49,028	49.7%	1,277	54.9%
Town	9,686	9.8%	353	10.1%
Rural	10,890	11.0%	374	12.5%
Total	98,557	100.0%	2,588	100.0%

Note. Population data from Illinois State Board of Education (2023) and Common Core of Data from the National Center for Education Statistics (2023a; 2023b). Educators in schools that did not report locale were excluded. Percentages are weighted. Details may not add up to total due to rounding.

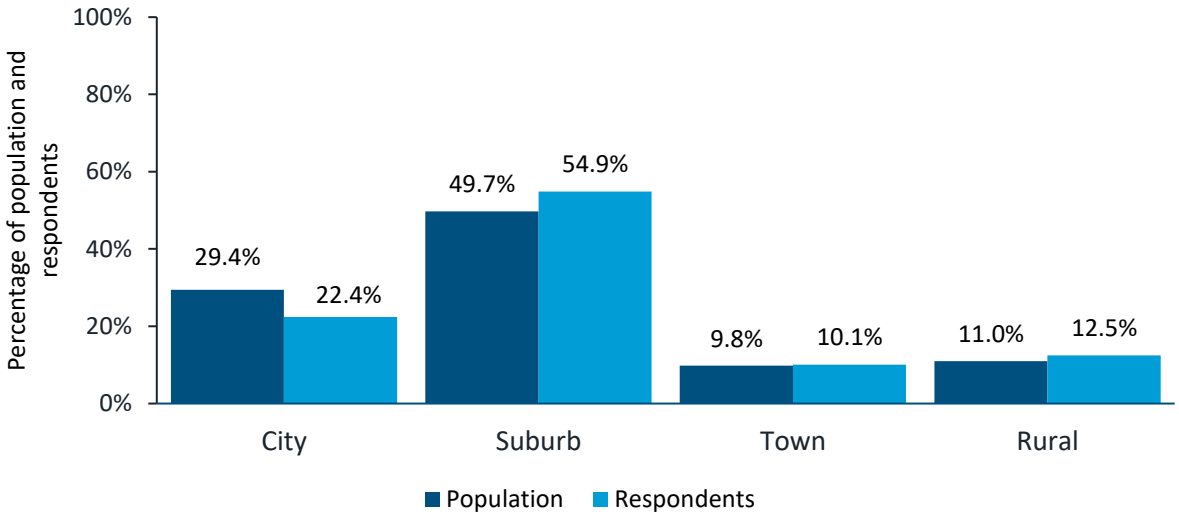
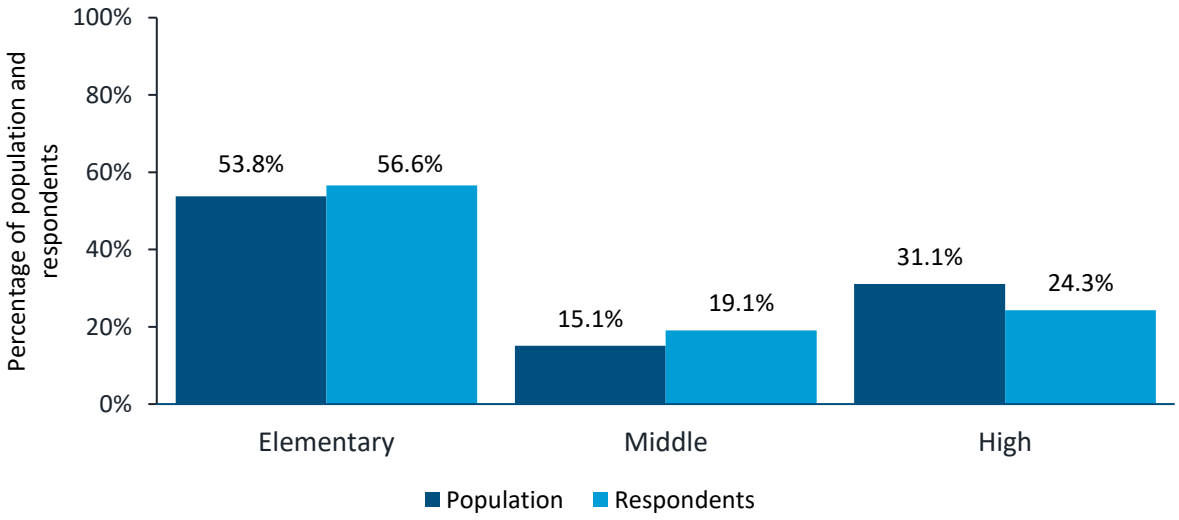


Exhibit 58 presents the distribution of school levels, as categorized in the Illinois Report Card data, of the schools in which the educator population and survey respondents worked in 2022–23. Approximately 56.6% of survey respondents worked in elementary schools, 19.1% in middle schools, and 24.3% in high schools.

Exhibit 58. School Levels: Population and Survey Respondents

School Level	Number in Population	Percent in Population	Number of Respondents	Percent of Respondents
Elementary	53,012	53.8%	1,176	56.6%
Middle	14,917	15.1%	339	19.1%
High	30,627	31.1%	1,073	24.3%
Total	98,557	100.0%	2,588	100.0%

Note. Population data from Illinois State Board of Education (2023). NA = Not available. Survey respondents who did not select a grade level were excluded. Percentage totals do not equal 100 because some educators worked with more than one grade level band. Percentages are weighted. Details may not add up to total due to rounding.



CPS enrolls 17.2% of students and employs 16.4% of teachers in the state. To better understand how the experiences of teachers in this unique district compared to the experiences of teachers in other districts, we disaggregate some findings into CPS and non-CPS schools.

Findings

This section presents the findings from the educator survey. First, key findings that address RQ1 to RQ5 are summarized. Then, findings for each RQ are described in more detail.

Key findings

RQ1: Components of District Evaluation Systems

- Over three quarters of survey respondents (76.5%) reported that their district's evaluation system used observation rubrics.

- Fifty-four percent of survey respondents reported that their evaluation system used districtwide or locally developed student assessments.
- More than one quarter of survey respondents (28.8%) reported that their district used standardized statewide assessments, and 8.4% reported that their district used other measures of educator practice.
- Higher proportions of respondents who worked in elementary (33.3%) and middle (33.1%) schools reported that standardized statewide assessments were used in their teacher evaluation systems compared to respondents who worked in high schools (16.2%).
- A lower proportion of Black or African American educators (36.2%) reported the use of districtwide assessments in their evaluation system compared to White educators (56.9%).
- Across demographic groups, high proportions of survey respondents (75.0% or more) reported the use of observation rubrics.
- Most educators reported feeling "Confident" in their understanding of their district's evaluation systems, with 52.1% feeling "Extremely Confident," 39.2% feeling "Somewhat Confident," and the remaining 8.7% feeling "Not Confident."

RQ2: Implementation Differences

- Overall, 97.6% of educators reported that they were observed at least once in the previous school year.
- Less than 3.0% of educators reported that they were never observed, 8.2% of educators reported that they did not engage in preobservation conferences, and 6.5% reported that they did not engage in postobservation conferences.
- A higher proportion of nontenured teachers reported that they were observed two or more times in the previous school year (97.5%) compared to tenured teachers (67.4%).
- A higher proportion of tenured teachers were never offered coaching (79.9%) compared to nontenured teachers (56.6%).

RQ3: Benefits of Evaluation System Components

- On the benefits of evaluation components, 31% to 39% of survey respondents found feedback, coaching, and the evaluation system "Extremely Actionable," "Extremely Useful," or "Extremely Supportive."
- Survey respondents' perceived usefulness of coaching varied on factors like race/ethnicity, years of experience, union affiliation, and school location.

- Survey respondents' reported levels of feedback actionability varied across demographic groups, work roles, and regions.

RQ4: Challenges in Implementation

- Some respondents perceived that the student assessment and formal observation components of the evaluation system measured their professional practice "Not Fairly at All."
- Overall, 16.3% of survey respondents reported that student assessments measured professional practice "Not Fairly at All" while 50.8% reported that formal observation reflected professional practice "Extremely Fairly."
- Respondents' perceived fairness of the evaluation system varied by demographic group, including race/ethnicity, union affiliation, and school locale.
- Respondents' perceptions of the accuracy of the evaluation system were similar across demographic groups.

RQ5: Stakeholder Suggestions for Addressing Challenges

- Most survey respondents (50.8%) found their evaluators to be "Extremely Knowledgeable."
- Getting support was considered "Not Difficult at All" by the majority (57.9%) of respondents.
- About one third (33.8%) of respondents reported no professional learning hours connected to evaluation feedback.
- Differences in perceptions of evaluator knowledge and support difficulty existed based on union affiliation and on school location, size, and locale.

RQ1. What are the components of districts' teacher and administrator evaluation systems?

To learn about the components of districts' educator evaluation systems, the survey prompted educators to select components that were included in their district's evaluation system. Observation rubrics for professional practice were widely adopted, with 76.6% of respondents reporting that their districts used observation rubrics (Exhibit 59), followed by districtwide or locally developed student assessments, which were reported by over half of respondents (54.0%). By contrast, only 28.8% of respondents reported that their districts used standardized statewide assessments, and 8.4% reported that their districts used other measures of educator practice.

Exhibit 59. Components of District Evaluation System

Which components did the district's evaluation system use to rate your professional practice during the previous school year (2022–23)? Please select all that apply:	Number of Respondents	Percent of Respondents (Standard Error)
Standardized, statewide assessments (e.g., Illinois Assessment of Readiness, Illinois Science Assessment, SAT)	659	28.8% (1.7)
Districtwide or locally developed student assessments, including student learning objectives	1,318	54.0% (2.2)
Observation rubrics for professional practice (e.g., Framework for Teaching or Classroom Assessment Scoring System)	1,783	76.6% (1.3)
Other measures of educator practice	206	8.4% (0.8)
Total	3,966	NA

Note. NA = Not available. Survey respondents who did not select a gender were excluded. The survey question was "Select all that apply."

To examine differences in responses by educator and school characteristics, cross tabulations of educators' selections of evaluation system components with educator demographic and professional characteristics, as well as school level, locale, and student composition were conducted. These results for each main evaluation system component are summarized below, and the corresponding tables are presented in Appendix F: Survey Results.

Observation Rubrics

The use of observation rubrics for professional practice was reported by similarly high proportions of educators across different demographic groups, such as race/ethnicity, gender, and experience levels (Exhibit F–1 through Exhibit F–4). One exception was that a higher proportion of nontenured teachers (81.1%) reported the use of observation rubrics for professional practice than administrators (66.8%) (Exhibit F–2).

Standardized Statewide Assessments

The use of standardized statewide assessments (i.e., Type I) varied across union affiliations (Exhibit F–3). Standardized statewide assessment utilization was reported by 31.9% of educators associated with the Illinois Education Association and by 29.8% of educators associated with the Illinois Federation of Teachers. These results were higher than what unaffiliated educators reported (13.4%), and the difference was statistically significant. In addition, standardized statewide assessment utilization was reported by 24.9% of educators associated with the Chicago Teachers Union and 19.9% of educators associated with "Other" unions, although the difference was not statistically significant. The percentage of respondents

who reported that standardized statewide assessments were used in their evaluation also varied across school levels (Exhibit F–4). Higher proportions of elementary school (33.3%) and middle school (31.1%) educators reported use of statewide assessments in their evaluations compared to high school educators (16.2%).

Districtwide and Local Assessments

The use of districtwide (i.e., Type II) or locally developed assessments (i.e., Type III) also varied with respect to the demographic and professional characteristics of respondents. A lower proportion of Black or African American educators reported use of districtwide or locally developed student assessments by their districts (36.2%) compared to White educators (56.9%) Exhibit 59). A lower proportion of related service providers reported use of districtwide or locally developed student assessments (22.9%) in their evaluation compared to over 50.0% for teachers and school administrators (Exhibit F–2).

In general, there was little difference in the use of evaluation components based on educator gender and role. The components that educators reported being used in their evaluation varied slightly by school service area, locale, whether they were in CPS, and school level, but they did not vary by school size, the percentage of students in the school from low-income families, or schoolwide ELA and mathematics proficiency rates.

Other Measures of Educator Practice

The proportion of survey respondents who reported use of other measures of educator practice was similar across educator and school characteristics. One exception was differences by work role. None of the related service providers reported use of other measures of educator practice, whereas 6.6% of nontenured teachers, 9.8% of tenured teachers, and 12.3% of school administrators reported use of other measures of educator practice in their evaluations (Exhibit F–2).

Educators' Confidence in Their Understanding of the District's Evaluation System

To learn about educators' understanding of their district's evaluation system, the survey asked respondents to report their level of confidence in their understanding. As shown in Exhibit 60, survey respondents reported feeling confident in their understanding of their district's evaluation system, with 52.1% feeling "Extremely Confident," and 39.2% feeling "Somewhat Confident." The remaining 8.7% reported feeling "Not Confident."

Exhibit 60. Confidence in Understanding Evaluation System

How confident did you feel in your understanding of the district's evaluation system during the previous school year (2022–23)?	Number of Respondents	Percentage of Respondents (Standard Error)
Not confident	172	8.7 (0.9)
Somewhat confident	897	39.2 (1.5)
Extremely confident	1,365	52.1 (1.5)
Total	2,434	100

Note. Educators who did not select a response or selected "Unsure" were excluded. Percentages are weighted. Details may not add up to total due to rounding.

Responses differed by educator and school characteristics (Exhibit F–5 through Exhibit F–7). There was a statistically significant difference between Black or African American (22.5%) and White educators (7.2%) in the proportion of survey respondents who reported feeling "Not Confident" in their understanding of the evaluation system (Exhibit F–5). A higher proportion of tenured teachers (56.8%) than nontenured teachers (46.4%) reported feeling "Extremely Confident" (Exhibit F–6). There was no statistically significant difference between nontenured teachers and administrators or between tenured teachers and administrators in their reported confidence in understanding the evaluation system. A lower proportion of CPS survey respondents (41.2%) reported feeling "Extremely Confident" than respondents in other school districts (54.2%) (Exhibit F–7).

RQ2. How do districts differ in their implementation of these components, and what factors (e.g., district needs, characteristics of the district) do stakeholders identify as potential reasons for these differences?

To understand the differences in district implementation of the evaluation system components, the survey asked educators about how often they were observed, how often they engaged in preconferences and postconferences related to an observation, how frequently they received written feedback based on an observation, and how frequently they were offered coaching as a result of an observation. This section reports the results for all survey respondents and by respondents' individual and school characteristics.

Overall, only a small percentage of survey respondents reported they were Never" observed (2.4%), "Never" engaged in preconferences (8.2%) or postconferences (6.5%), or "Never" received written feedback (5.6%). As shown Exhibit 61, 79.4% of respondents reported being observed "Two or More Times" during the previous school year. About half of educators reported being engaged in preconferences (50.4%, Exhibit 62) or postconferences (52.9%,

Exhibit 63) “Two or More Times.” A significant majority of educators (64.0%) reported receiving written feedback based on an observation “Two or More Times” during the previous school year (Exhibit 64). As shown in Exhibit 65, 69.9% of educators reported “Never” receiving coaching as a result of an observation, 12.5% reported receiving coaching “Once,” and 17.5% reported receiving coaching “Two or More Times.”

Exhibit 61. Observation Frequency in School Year 2022–23

How frequently was your professional practice observed as part of your district’s evaluation system during the previous school year (2022–23)?	Number of Respondents	Percentage of Respondents (Standard Error)
Never	56	2.4 (0.5)
Once	563	18.2 (1.7)
Two or More Times	1,916	79.4 (1.9)
Total	2,535	100

Note. Educators who did not select a response or selected “Unsure” were excluded. Percentages are weighted. Details may not add up to total due to rounding.

Exhibit 62. Preconference Frequency in School Year 2022–23

How frequently did you engage in a preconference before you were observed as part of your district’s evaluation system during the previous school year (2022–23)?	Number of Respondents	Percentage of Respondents (Standard Error)
Never	209	8.2 (0.9)
Once	1,114	41.4 (1.6)
Two or More Times	1,235	50.4 (1.5)
Total	2,558	100

Note. Educators who did not select a response or selected “Unsure” were excluded. Percentages are weighted. Details may not add up to total due to rounding.

Exhibit 63. Postconference Frequency in School Year 2022–23

How frequently did you engage in a postconference after you were observed as part of your district's evaluation system during the previous school year (2022–23)?	Number of Respondents	Percentage of Respondents (Standard Error)
Never	140	6.5 (0.9)
Once	1,068	40.6 (1.8)
Two or More Times	1,346	52.9 (1.8)
Total	2,554	100

Note. Educators who did not select a response or selected "Unsure" were excluded. Percentages are weighted. Details may not add up to total due to rounding.

Exhibit 64. Frequency of Receiving Written Feedback in School Year 2022–23

How frequently did you receive written feedback based on an observation as part of your district's evaluation system during the previous school year (2022–23)?	Number of Respondents	Percentage of Respondents (Standard Error)
Never	120	5.6 (0.9)
Once	836	30.4 (2.1)
Two or More Times	1,594	64.0 (2.5)
Total	2,550	100

Note. Educators who did not select a response or selected "Unsure" were excluded. Percentages are weighted. Details may not add up to total due to rounding.

Exhibit 65. Frequency of Being Offered Coaching in School Year 2022–23

How frequently were you offered coaching as a result of an observation as part of your district's evaluation system during the previous school year (2022–23)?	Number of Respondents	Percentage of Respondents (Standard Error)
Never	1,570	69.9 (1.7)
Once	288	12.5 (1.0)
Two or More Times	400	17.5 (1.5)
Total	2,258	100

Note. Educators who did not select a response or selected "Unsure" were excluded. Percentages are weighted. Details may not add up to total due to rounding.

To examine variations in implementation differences by educator and school characteristics, cross tabulations of the responses to questions about observations, pre- and postconferences, written feedback, and coaching with educator demographic and professional characteristics as well as school level, locale, and student composition were conducted. Results from these analyses are summarized below. The corresponding tables are presented in Appendix F Exhibit F–8 through Exhibit F–23.

Observation Frequency

During the previous school year (2022–23), most educators (97.6%) reported being observed at least once as part of the district's evaluation system. There were a few notable differences among subgroups of survey respondents by race/ethnicity, tenured status, experience, and educational region of their school.

Black or African American (90.0%) and Hispanic or Latinx (90.9%) educators were more likely to report being observed "Two or More Times" compared to White educators (77.4%) (Exhibit F–8). Nontenured teachers were observed more frequently ("Two or More Times": 97.5%) compared to tenured teachers ("Two or More Times": 67.4%) (Exhibit F–9). Educators in their work role for 6 years or more were less likely (65.4%) to be observed two or more times compared to educators with less experience (Exhibit F–10).

Frequency of observation varied by the educational region of the school. Educators in Region 4 (84.8%) and Region 5 (86.8%) reported a higher percentage of being observed "Two or More Times" compared to educators in Region 2 (69.0%) (Exhibit F–11).

Frequency of Preconferences

Among all survey respondents, approximately 50% of educators reported being engaged in a preconference before an observation "Two or More Times." Reported frequency of preconferences differed by work role, experience, tenure status, and whether or not educators worked in CPS schools.

As shown in Exhibit F–12, nontenured teachers reported a higher percentage (86.4%) of having a preconference "Two or More Times" than tenured teachers (25.2%), school administrators (34.6%), and related service providers (52.0%). Conversely, nontenured teachers (2.4%) were less likely to report "Never" having a preconference than tenured teachers (9.9%) and administrators (28.4%); and nontenured teachers (11.2%) were less likely to report having a preconference "Once" than tenured teachers (64.8%), administrators (37.0%), and related service providers (39.2%).

Survey respondents in schools for 6 years or more were less likely (24.0%) to report being engaged in a preconference "Two or More Times" compared to educators with less experience (80.5% of educators in their first year, 77.3% of educators in their second or third year, and 50.9% of teachers in their fourth or fifth year at the school) (Exhibit F–13). Educators associated with the Illinois Education Association (51.9%) and other unions (61.1%) reported higher percentages of engaging in preconferences "Two or More Times" compared to educators associated with the Illinois Federation of Teachers (42.0%) (Exhibit F–14). Educators in CPS schools (52.5%) were more likely to report being engaged "Once" in a preconference than educators in non-CPS schools (39.2%) (Exhibit F–15).

Frequency of Postconferences

Similar patterns are shown in postconference engagement, with differences in frequencies reported by educators' tenure status, work role, experience, and union affiliation.

Nontenured teachers had a higher percentage (86.1%) of engaging in a postconference "Two or More Times" than tenured teachers (29.4%), school administrators (44.7%), and related service providers (57.6%) (Exhibit F–16). Conversely, tenured teachers (62.8%) were more likely to report having a preconference "Once" than administrators (36.3%).

Educators in schools for 6 years or more had a lower percentage (27.7%) of engaging in a postconference "Two or More Times" compared to educators with less experience (82.6% of educators in their first year, 76.6% of educators in their second or third year, and 54.7% of educators at their fourth or fifth year at the school) (Exhibit F–17). With respect to union affiliation, unaffiliated educators had a lower percentage (38.0%) of engaging in a postconference "Two or More Times" compared to educators associated with other unions (67.7%); educators associated with the Illinois Federation of Teachers had a lower percentage (45.3%) of engaging in a postconference "Two or More Times" compared to educators associated with the Illinois Education Association (54.9%) or other unions (67.7%). In addition, educators associated with the Chicago Teachers Union had the lowest percentage (43.0%) of engaging in a postconference "Two or More Times" (Exhibit F–18), although the difference was not statistically significant.

Frequency of Receiving Written Feedback

Survey respondents reported different frequencies of receiving written feedback based on an observation by tenure status, years of experience, and whether or not they worked in CPS schools.

A higher percentage of nontenured teachers reported receiving written feedback "Two or More times" (86.8%) than tenured teachers (48.2%) (Exhibit F–19). As shown in Exhibit F–20, a lower

percentage (46.1%) of educators in their role for 6 years or more reported receiving written feedback "Two or More Times" compared to educators with less experience (84.4% of educators in their first year, 80.4% of educators in their second or third year, and 66.5% of educators at their fourth or fifth year the school). A lower percentage of educators in their role for 4 or 5 years (46.1%) reported receiving written feedback "Two or More Times" compared to educators with 1 year of experience or less (84.4%). A lower percentage of CPS educators (47.4%) reported receiving written feedback "Two or More Times" than educators in non-CPS schools (67.2%, Exhibit F–21).

Frequency of Being Offered Coaching

Survey responses showed differences in the frequencies of being offered coaching by educators' work role and years of experience. A higher percentage of tenured teachers (79.9%) reported "Never" being offered coaching than nontenured teachers (56.6%) (Exhibit F–22). A lower percentage (50.5%) of educators in their role for the first year reported "Never" being offered coaching compared to educators with more experience (63.9% of educators in their second or third year, 71.8% of educators in their fourth or fifth year, and 80.8% of educators with more than five years of experience at their school) (Exhibit F–23).

RQ3. What benefits have stakeholders experienced implementing components of the teacher and administrator evaluation system with respect to the following: obtaining useful feedback, supporting improvements to professional practice, informing policy decisions, and adapting to the COVID-19 pandemic?

To address this RQ, the survey asked educators about the actionability of written and verbal feedback they received, the usefulness of coaching, and the perceived supportiveness of the district's evaluation system. Results for all survey respondents and by respondent characteristics are reported below.

As presented in Exhibit 66, 89.3% of survey respondents reported that the written and verbal feedback they received as part of their evaluation system was "Somewhat Actionable" or "Extremely Actionable." The remainder (10.6%) of educators indicated that feedback was "Not at All Actionable."

Exhibit 66. Actionability of Written and Verbal Postconference Feedback

How actionable was the written and verbal postconference feedback that you received as part of your district's evaluation system during the previous school year (2022–23)?	Number of Respondents	Percentage of Respondents (Standard Error)
Not at all actionable	229	10.6 (1.1)

How actionable was the written and verbal postconference feedback that you received as part of your district's evaluation system during the previous school year (2022–23)?	Number of Respondents	Percentage of Respondents (Standard Error)
Somewhat actionable	1,210	55.9 (1.6)
Extremely actionable	774	33.4 (1.5)
Total	2,213	100

Note. Educators who did not select a response or selected "Unsure" were excluded. Percentages are weighted. Details may not add up to total due to rounding.

Exhibit 67 reports the usefulness of coaching received as part of the evaluation system. A large majority (88.6%) of survey respondents reported that the coaching they received was "Somewhat Useful" or "Extremely useful"; the remaining 11.5% stated that coaching was "Not at All Useful."

Exhibit 67. Usefulness of Coaching

How useful was the coaching you received as part of your district's evaluation system during the previous school year (2022–23)?	Number of Respondents	Percentage of Respondents (Standard Error)
Not at all useful	46	11.5 (2.0)
Somewhat useful	344	50.1 (2.8)
Extremely useful	253	38.5 (2.3)
Total	643	100

Note. Educators who did not select a response or selected "Unsure" were excluded. Percentages are weighted. Details may not add up to total due to rounding.

Exhibit 68 presents the results for the supportiveness of the evaluation system to the educators' professional growth and development. Overall, 79.8% of survey respondents reported that their district's evaluation system was "Somewhat Supportive" or "Extremely Supportive" while 20.2% reported that their district's evaluation system was "Not at All Supportive" for their professional growth and development.

Exhibit 68. Supportive Evaluation System

Overall, how supportive was the district's evaluation system to your professional growth and development during the previous school year (2022–23)?	Number of Respondents	Percentage of Respondents (Standard Error)
Not at all supportive	419	20.2 (1.6)
Somewhat supportive	1,173	49.0 (1.3)
Extremely supportive	808	30.8 (1.5)
Total	2,400	100

Note. Educators who did not select a response or selected "Unsure" were excluded. Percentages are weighted. Details may not add up to total due to rounding.

In addition to these analyses for all survey respondents, cross tabulations to examine differences in survey responses by educator and school characteristics were conducted. The findings are summarized in the following sections and reported in Appendix F Exhibit F–24 through Exhibit F–31.

Actionability of Feedback

Survey respondents reported different levels of actionability on the feedback they received based on their role, school region and locale, and whether they worked in CPS schools.

Higher proportions of related service providers (62.7%) and nontenured teachers (38.3%) reported their feedback as "Extremely Actionable" compared to tenured teachers (28.3%) (Exhibit F–24). Higher proportions of educators in Region 6 (56.6%) reported receiving "Extremely Actionable" feedback compared to educators in other regions (under 40.0%) (Exhibit F–25). Compared to educators in city schools (24.0%), a higher proportion of educators in other areas reported receiving "Extremely Actionable" feedback (35.3% for suburb; 41.0% for town; 35.7% for rural; Exhibit F–26). The difference between educators in CPS and non-CPS schools was also significant: the percentage of educators in CPS schools who reported receiving "Extremely Actionable" feedback was 16.6 percentage points lower than educators in other districts (Exhibit F–27).

Usefulness of Coaching

Survey respondents reported different levels of usefulness for the coaching they received by experience. A higher proportion of educators in their first year (49.4%) reported their coaching was "Extremely Useful" compared to educators with 4 or 5 years of experience (25.8%) or with 6 years or more (28.0%) (Exhibit F–28).

Supportiveness of District's Evaluation System for Professional Growth and Development

Regarding the supportiveness of the district's evaluation system for professional growth and development, there were differences in survey responses by educators' role, whether they worked in CPS, and the school's ELA achievement proficiency quartile.

A higher proportion of nontenured teachers (37.3%) reported that their district's evaluation system was "Extremely Supportive" than tenured teachers (25.8%) and administrators (24.0%) (Exhibit F–29). The percentage of CPS educators who reported the evaluation system was "Extremely Supportive" (19.4%) was lower than non-CPS educators (32.9%) (Exhibit F–30). A lower proportion (25.6%) of educators in schools with ELA proficiency rates in the 1st (lowest) quartile reported their district's evaluation system was "Extremely Supportive" than educators in schools in higher ELA proficiency quartiles (34.4% of educators in schools in the 2nd quartile, 31.9% of educators in schools in the 3rd quartile, and 35.0% of educators in schools in the highest quartile of ELA proficiency rates) (Exhibit F–31).

RQ4. What challenges have stakeholders experienced while implementing components of the teacher and administrator evaluation system with respect to the following: obtaining useful feedback, supporting improvements to professional practice, informing policy decisions, reported biases or inequities in system components, and adapting to the COVID-19 pandemic?

To identify challenges stakeholders experienced with the evaluation system, the survey asked educators about the fairness of its student assessment and formal observation components and their accuracy in measuring the effectiveness of the educator's overall professional practice.[14] Results are reported below for all survey respondents and by educator and school characteristics.

As shown in Exhibit 69, overall, 34.9% of respondents reported that the student assessment component measured their professional practice "Extremely Fairly," while 48.8% thought the component measured "Somewhat Fairly," and 16.3% believed the component measured "Not Fairly at All."

[14] As a reminder, the survey used "student assessment component" as a synonym for the "student growth component" and "formal observation component" as a synonym for "professional practice component." Because survey respondents might not have considered these to be synonyms, we report findings based on the language in the survey.

Exhibit 69. Student Assessment Component Fairness

How fairly did the student assessment component of the district's evaluation system measure your professional practice during the previous school year (2022–23)?	Number of Respondents	Percentage of Respondents (Standard Error)
Not fairly at all	210	16.3 (1.5)
Somewhat fairly	722	48.8 (2.0)
Extremely fairly	553	34.9 (2.1)
Total	1,485	100

Note. Educators who did not select a response or selected "Unsure" were excluded. Percentages are weighted. Details may not add up to total due to rounding.

Exhibit 70 shows the overall results for the perceived fairness of the formal observation component. Slightly more than half (50.8%) of respondents reported that the formal observation measured their professional practice "Extremely Fairly" while 42.0% thought the component measured "Somewhat Fairly." Only 7.2% believed the formal observation component measured their professional practice "Not Fairly at All."

Exhibit 70. Formal Observation Component Fairness

How fairly did the formal observation component of the district's evaluation system measure your professional practice during the previous school year (2022–23)?	Number of Respondents	Percentage of Respondents (Standard Error)
Not fairly at all	100	7.2 (1.0)
Somewhat fairly	667	42.0 (1.9)
Extremely fairly	957	50.8 (1.9)
Total	1,724	100

Note. Educators who did not select a response or selected "Unsure" were excluded. Percentages are weighted. Details may not add up to total due to rounding.

Exhibit 71 shows the results for the perceived accuracy of the district's evaluation system in measuring the effectiveness of the educator's overall professional practice. More than one third (39.5%) of respondents reported the evaluation system measured "Extremely accurately," 50.0% thought their practice was measured "Somewhat Accurately," and 10.5% reported "Not at All Accurately."

Exhibit 71. Accuracy of Evaluation System

How accurately did the district's evaluation system measure the effectiveness of your overall professional practice during the previous school year (2022–23)?	Number of Respondents	Percentage of Respondents (Standard Error)
Not at all accurately	200	10.5 (0.9)
Somewhat accurately	1,113	50.0 (1.4)
Extremely accurately	947	39.5 (1.5)
Total	2,260	100

Note. Educators who did not select a response or selected "Unsure" were excluded. Percentages are weighted. Details may not add up to total due to rounding.

In addition to the overall results reported above, cross tabulations to examine differences in responses to survey questions about the evaluation system's fairness and accuracy by educator and school characteristics were conducted. Notable differences are summarized below, and the tables reporting all the cross tabulation results are presented in Appendix F, Exhibit F–32 through Exhibit F–46.

Fairness of Student Assessment Component in District Evaluation

Responses to the survey question on the fairness of the student assessment component varied by educator and school characteristics, such as subject taught by the educator, school region, locale, CPS school versus non-CPS school, and the school's achievement proficiency quartiles in ELA and math.

A lower proportion of general education teachers (28.0%) reported that the student assessment component measured their professional practice "Extremely Fairly" than elective subject teachers (42.8%) (Exhibit F–32). A higher proportion of educators in Region 1 (19.1%) reported that their performance was measured by the student assessment component "Not Fairly at All" than those in Region 4 (9.1%) (Exhibit F–33). The percentage of educators in city schools (22.5%) who reported they were measured "Extremely Fairly" by the student assessment component was lower compared to educators in rural schools (43.7%) (Exhibit F–34). The percentage of CPS educators (28.4%) who reported the student assessment component measured their professional practice "Not Fairly at All" was higher than educators in other districts (14.4%) (Exhibit F–35). A higher proportion of educators in schools in the 1st (lowest) quartile of ELA (20.2%) and mathematics (20.4%) proficiency rates reported that the student assessment component measured their performance "Not Fairly at All" than educators

in schools in the 4th (highest) quartile of ELA (11.4%) and mathematics (12.2%) proficiency rates (Exhibit F–36 and Exhibit F–37).

Fairness of Formal Observation Component in District Evaluation

Perceptions of the fairness of the formal observation component differed by educator and school characteristics. A higher percentage of Black or African American educators (23.8%) than White educators (5.9%) reported that the formal observation component measured their professional practice "Not Fairly at All," and a lower percentage of Black or African American educators (30.6%) than White educators (53.1%) reported the component measured "Extremely Fairly" (Exhibit F–38).

Higher proportions of educators affiliated with the Illinois Education Association (57.2%) and Illinois Federation of Teachers (52.6%) reported that the formal observation component measured their professional practice "Extremely Fairly" than those affiliated with the Chicago Teachers Union (31.7%). In addition, the formal observation component was reported as "Extremely Fairly" by 51.1% of educators associated with the other unions and 38.6% of unaffiliated educators (Exhibit F–39), although the difference was not statistically significant. A lower percentage of educators in city schools (35.3%) reported that the formal observation component measured "Extremely Fairly" compared to teachers in schools in other locales (Exhibit F–40). Lastly, the percentage of CPS educators (30.2%) who reported that the observation component measured "Extremely Fairly" was lower than the percentage of educators in other districts (55.2%) (Exhibit F–41).

Accuracy of District Evaluation of Professional Practice

Responses to the survey question about the accuracy of the district's evaluation system also differed by educator race/ethnicity, union affiliation, and school characteristics, such as locale and the percentage of low-income students enrolled.

A lower percentage of Black or African American educators (24.7%) reported that the system measured their professional practice "Extremely Accurately" compared to 41.4% of White educators (Exhibit F–42). A higher percentage of educators associated with the Chicago Teachers Union (21.6%) reported that the system measured their professional practice "Not at All Accurately" than those associated with the Illinois Education Association (8.1%) and Illinois Federation of Teachers (9.8%) (F–43).

The percentage of educators in city schools who reported the system measured their professional practice "Not at All Accurately" (18.8%) was higher than educators in other locales (F–44). A lower percentage of CPS educators (20.3%) reported the system measured their professional practice "Not at All Accurately" than educators in other districts (8.6%) (Exhibit F–

45). Lastly, a higher percentage of educators (44.2%) in schools in the 1st (lowest) quartile for the enrollment of students from low-income families reported that the system measured their professional practice "Extremely Accurately" than educators in schools in the 4th (highest) low-income quartile (32.2%) (Exhibit F–46).

RQ5. How do stakeholders suggest addressing reported challenges? What supports, resources, and changes are recommended?

To address this RQ, the survey asked educators about how knowledgeable their evaluator was about the educator's professional practice, the number of hours of professional learning the educator received, and the difficulty of getting support to improve practice. In this section, the overall responses to these questions and responses by educator and school characteristics are reported.

Exhibit 72 shows survey respondents' perception of how knowledgeable their evaluators were about their professional practice. Slightly more than half (50.8%) of survey respondents reported that their evaluator was "Extremely Knowledgeable" and 40.4% reported "Somewhat Knowledgeable."

Exhibit 72. Perception of Evaluator Knowledge

How knowledgeable was your evaluator about your professional practice during the previous school year (2022–23)?	Number of Respondents	Percentage of Respondents (Standard Error)
Not at all knowledgeable	179	8.7 (1.0)
Somewhat knowledgeable	884	40.4 (1.4)
Extremely knowledgeable	1,244	50.8 (1.8)
Total	2,307	100

Note. Educators who did not select a response or selected "Unsure" were excluded. Percentages are weighted. Details may not add up to total due to rounding.

Exhibit 73 shows the number of professional learning hours the survey respondents reported receiving. More than half (52.6%) of respondents reported receiving no professional learning ("None") or "Less Than 1 Hour"; 33.7% reported receiving between 1 and 5 hours, and 13.7% reported receiving 6 hours or more.

Exhibit 73. Hours of Professional Learning

Think of the hours of professional learning you accessed or received during the previous school year (2022–23). How many of these hours, if any, were directly connected to feedback you received as part of your evaluation?	Number of Respondents	Percentage of Respondents (Standard Error)
None	669	33.8 (1.8)
Less than 1 hour	406	18.8 (1.3)
1–2 hours	427	21.2 (1.2)
3–5 hours	290	12.5 (1.0)
6–10 hours	132	6.8 (0.7)
11 or more hours	134	6.9 (0.8)
Total	2,058	100

Note. Educators who did not select a response or selected "Unsure" were excluded. Percentages are weighted. Details may not add up to total due to rounding.

Exhibit 74 presents survey respondents' reported perceived difficulty in accessing support when to improve their practice. A majority (57.9%) reported it was "Not at All Difficult" to get support, 27.7% reported it was "Somewhat Difficult," and 14.4% reported that getting support was "Extremely Difficult."

Exhibit 74. Difficulty of Getting Support

If you requested extra support (e.g., informal observations, informal feedback, or informal mentorship from colleagues) to improve your practice, how difficult was it to get the support that you needed during the previous school year (2022–23)?	Number of Respondents	Percentage of Respondents (Standard Error)
Not at all difficult	623	57.9 (2.6)
Somewhat difficult	229	27.7 (2.1)
Extremely difficult	92	14.4 (1.5)
Total	944	100

Note. Educators who did not select a response or selected "Unsure" were excluded. Percentages are weighted. Details may not add up to total due to rounding.

Educators' survey responses on how knowledgeable their evaluators were about their professional practice, the number of professional learning hours they received, and how difficult it was to get support differed by educator and school characteristics. The cross

tabulation results are presented below, and detailed findings are located in Appendix F, Exhibit F–47 through Exhibit F–54.

Knowledge of Evaluator

Survey data suggested differences in reported evaluator knowledge across various groups of educators and schools. There was no difference across various union representations in the percentage of educators who reported their evaluators as "Not at All Knowledgeable." However, a lower percentage of unaffiliated educators (28.4%) reported that their evaluators were "Extremely Knowledgeable" compared to educators associated with the Illinois Education Association (53.3%) and Illinois Federation of Teachers (53.4%). In addition, 43.5% of educators associated with the Chicago Teachers Union and 52.8% of educators associated with Other unions reported that their evaluators were "Extremely Knowledgeable" (Exhibit F–47), although the difference was not statistically significant.

When data were disaggregated by school locale, survey data showed a higher percentage of educators in city schools (12.5%) reporting that their evaluators were "Not at All Knowledgeable" than educators in town areas (3.7%); by contrast, a higher percentage of educators in rural areas (60.8%) reported that their evaluators were "Extremely Knowledgeable" than those in cities (43.8%) (Exhibit F–48).

Hours of Professional Learning Received

The reported number of professional learning hours connected to the evaluation also varied across groups of teachers and schools. A higher proportion of female educators (35.9%) reported having received no hours of professional learning ("None") compared to their male counterparts (26.7%) (Exhibit F–49).[15] A higher percentage of educators with 1 year of experience reported receiving "11 or More Hours" (12.6%) of professional learning than those with 6 or more years of experience (4.5%) (Exhibit F–50).

Difficulty in Obtaining Needed Support

The reported difficulty in obtaining support for improving professional practice also varied by educator and school characteristics. A higher percentage of educators affiliated with the Chicago Teachers Union (27.1%) reported that receiving support was "Extremely Difficult" than educators associated with the Illinois Education Association (10.4%) and the Illinois Federation of Teachers (11.5%). Percentages of educators associated with Other unions and Unaffiliated educators who reported that receiving support was "Extremely Difficult" were suppressed because fewer than 10 respondents reported it. Chicago Teachers Union respondents also were

[15] Results for other categories of gender are not reported to protect respondent privacy (U.S. Department of Education, 2023; Seastrom, 2010).

the lowest percentage of educators who reported that obtaining support was "Not at All Difficult" (32.0%), as compared to educators associated with the Illinois Education Association (61.3%), the Illinois Federation of Teachers (61.4%), and other unions (69.6%). The percentage of unaffiliated educators who reported that receiving support was "Not at All Difficult" was suppressed because fewer than 10 respondents reported it (Exhibit F–51).

By locale, a higher percentage of educators in city schools (27.8%) reported getting support was "Extremely Difficult" than educators in schools in other locales (11.6% of educators in suburban schools and 9.0% of educators in towns),[16] and a lower percentage of educators in cities (37.2%) reported getting support was "Not at All Difficult" compared to educators in other locales (64.4% of educators in suburban schools, 59.8% of educators in schools in towns, and 71.0% of educators in rural areas) (Exhibit F–52). A lower percentage of CPS educators (29.1%) reported that getting support was "Not at All Difficult" than educators in other districts (64.7%). The percentage of CPS educators who reported getting support was "Extremely Difficult (32.5%) was also higher than educators in other districts (10.1%) (Exhibit F–53). Almost one fourth (22.4%) of educators in schools in the 1st (lowest) quartile of ELA achievement proficiency reported that getting support was "Extremely Difficult"; by contrast, 8.9% and 9.2% of educators in schools in the third and 4th (highest) quartiles, respectively, reported that getting support was "Extremely Difficult" (Exhibit F–54).

[16] Statistics are not reported for rural schools because the group being analyzed is too small.

Interview and Focus Group Analysis

Introduction and Purpose

To gain a deeper, nuanced understanding of PERA implementation, interviews and focus groups with a range of educators and policy makers throughout Illinois were conducted. The approach to qualitative data collection was organized as follows:

- Twelve interviews with early contributors to PERA legislation and implementation. This group included original members of PEAC, contributors to the writing of the legislation, and others who were part of the early negotiations and implementation processes.
- Thirteen focus groups with PERA joint committee members from multiple districts.
- Three focus groups with PK–12 school administrators and seven focus groups with teachers from districts across Illinois who experienced the evaluation system first-hand as either evaluators or the recipients of one or more evaluations.

Interviews and focus group protocols were developed in a partnership of PEAC and ISBE (Appendices G–K). These protocols delved into multiple facets of the research questions, including the evaluation system's components as well as the benefits and challenges that participants experienced. Interview and focus group participants offered valuable insight into how state-, district-, and school-level systems interact with and inform educators' experiences with the evaluation system in Illinois. Interviews with early contributors also provided perspectives on the historical context leading up to PERA's passage and enactment, how the implementation has unfolded since PERA's passage, and the statewide implications of PERA implementation.

Broadly, the use of qualitative data to identify potential explanations and nuances in quantitative data is grounded in established mixed methods research designs (e.g., Czarniawska, 2008; Ivankova et al., 2006; Maxwell, 2021; Patton, 2002). Situated within this study, analysis of interview and focus group data serves dual purposes. First, this work helps to explain and elaborate on several patterns found in the survey data. Second, interviews and focus groups help in identifying nuances within these broad patterns, thereby shedding light on variations in educators' experiences as well as how those experiences are nested within broader systems. In the subsections that follow, methods for and findings from gathering and analyzing data from interviews and focus groups are shared.

Interviews with Early Contributors

A list of 22 early contributors to PERA, including founding members of PEAC, PERA legislators and policy writers, and leaders of professional organizations who were involved with PERA's implementation at the state level (e.g., representatives from Illinois Federation of Teachers, Illinois Education Association, Illinois Principals Association), was compiled, and these individuals were invited to share their experiences and insights in a 1-hour virtual interview. The central foci of these interviews were participants' roles in the process, understanding the goals and purpose of the legislation, and perceived strengths and challenges facing those responsible for its implementation. Twelve early contributors opted to participate in an interview during fall/winter 2023.

Focus Groups With PERA Joint Committee Members

In addition to interviews of early contributors, 13 focus groups with district joint committee members were conducted in fall/winter 2023–24 to learn about their experiences with and perspectives on PERA implementation.[17]

Specifically, a subset of joint committees across the state were invited to participate in a focus group to gain insights into how districts interpreted PERA and developed their evaluation plans, and to learn whether and how districts made changes to these plans over time. AIR drew a stratified random subset of 24 districts from the larger survey sample of 125 districts. Strata were formed based on two variables, geographic region (six regions) and percentage of students in the district who are eligible for FRPL (high or low[18]), thereby forming 12 strata (e.g., Region 1 × Low FRPL, Region 1 × High FRPL, etc.). Within the 12 strata, districts were sorted by key characteristics, including locale (city, suburban, town/rural), student enrollment, student demographic composition (i.e., percentages of White, Black or African American, and Hispanic or Latinx students), and number of teachers. To ensure a comprehensive collection of voices, two districts from each of the 12 strata were selected, forming a sample of 24 districts.

Anticipating that not all districts would respond to the opportunity to provide PERA joint committee members' names and emails—and based on literature regarding data saturation in qualitative research (Guest et al., 2006)—the research team agreed that focus group participation from either district within each of the 12 strata was sufficient. In other words, once a response from one of the district pairs in our subsample was received, the other district

[17] ISBE's nonregulatory guidance (2015) states that all districts should establish a joint committee, composed of an equal number of district representatives and teachers, to develop the structure of the district's evaluation plan (i.e., observations of professional practice; indicators of student growth; etc.). The Illinois General Assembly Compiled Statutes (2024) also recommend that joint committees meet at least annually to "assess and review the effectiveness of the district's evaluation plan for the purposes of continuous improvement of instruction and evaluation practices."

[18] "High" refers to districts above the median value listed in **Error! Reference source not found.Error! Reference source not found.**, and "Low" refers to districts below the median value.

was no longer invited to participate in a focus group. Note that, although CPS was not part of the systematically selected joint committee subsample, it was determined that CPS should receive an invitation to participate, given that it is the largest district in the state. Members of CPS's joint committee agreed to participate in the focus groups.

In sum, the research team conducted focus groups with joint committees from 12 of the 24 randomly sampled districts, plus CPS, to aid in developing a comprehensive data set for analysis and identifying recommendations. Of these 13 districts, four were in cities, three were suburban, and six were town/rural. There was roughly an even split between high- and low-income districts, and two or three districts from each of Service Areas 1–5 participated. Despite outreach and recruitment efforts, only one district from Service Area 6 participated.

Focus Groups with Teachers and School Administrators

In January and February 2024, 10 focus groups with teachers and school-based administrators were conducted to learn about their experiences with their district's evaluation systems. These individuals were selected based on their survey participation. Specifically, the final item on the survey asked teachers and school-based administrators if they would be interested in volunteering for a focus group. Of the 2,588 survey respondents, 229 teachers and 33 school-based administrators volunteered. However, not all volunteers were able to participate due to scheduling constraints; in total, 36 teachers and 16 administrators participated in focus groups.

To form teacher focus groups, participants were grouped according to district locale (i.e., city, suburban, and town/rural) and FRPL status (i.e., high-income, low-income), resulting in six different focus group compositions (e.g., Suburban x High-income, City x Low-income, etc.). To form focus groups with school-based administrators, participants were grouped solely by district locale (i.e., city, suburban, town, rural), because far fewer administrators expressed interest in participating in focus groups.

For additional information on participant outreach and informed consent, please see Appendix L.

Data Collection

Semistructured interview and focus group protocols designed to elicit in-depth responses to relevant research questions were developed based on feedback from PEAC and ISBE. After protocols were finalized, early contributors to PERA were invited to participate in interviews, and members of PERA joint committees were invited to participate in focus groups to provide insights into the early days of PERA implementation, including district-level contexts of implementation and efforts to tailor PERA to local situations. Twelve early contributors to PERA

were interviewed between October and December 2023, and between November 2023 and February 2024, 13 focus groups were conducted with PERA joint committee members.

In December 2023, the research team recruited teacher and administrator focus group participants, drawing on the sample of survey participants who indicated "Yes" to participate in these focus groups. Seven focus groups with 36 teachers and three focus groups with 16 administrators occurred between January and February 2024.[19] Each focus group included no more than eight participants to enable the voices and perspectives of all participants to be heard.

All interviews and focus groups were conducted virtually on Zoom. Verbal consent to record these conversations was obtained. Interviews and focus groups took between 30 and 75 minutes and were scheduled at times that were respectful of participants' work responsibilities (e.g., after school for teachers and joint committee members).

Data Analysis

Focus group and interview transcripts were analyzed to address RQs 1–5. A coding scheme was designed using NVivo analytic software, and the research team coded each data source based on that coding scheme. Researchers analyzing data underwent training to ensure intercoder reliability (i.e., agreement across coders).

Analysis of interview and focus group responses included a multistep analytical process to ensure that the findings are defensible and transparent (i.e., able to be demonstrated to others; Thomas, 2006). Content of interviews and focus groups was analyzed deductively and inductively. Through deductive analysis, predetermined codes were applied to the data, meaning excerpts of data were "tagged" with codes corresponding to the study's RQs. After this initial coding process, inductive codes were developed and applied. Inductive analysis entails examining the data holistically to identify emergent themes, enabling the uncovering of thematic patterns beyond those anticipated based on the research questions. The findings of the deductive and inductive analyses were combined to compare patterns and themes from within and across participant type (i.e., early contributor, joint committee member, administrator, teacher) and district characteristics (e.g., locale and high- and low-income status). Throughout this process, the research team systematically categorized the data through reduction, organization, and connection (Charmaz, 2008; Denzin & Lincoln, 2003; Dey, 1993; LeCompte, 2000; Maxwell, 2013).

[19] Three additional, alternative focus group times for teachers unable to make the originally scheduled group time were proposed. The suburban and town/rural alternative groups did not have any attendees. The urban alternative group consisted of teachers from urban, low-FRPL districts only.

In Appendix M, the processes the research team followed to develop and apply coding, establish intercoder reliability, and engage in subsequent rounds of analysis after initial coding are described. The finalized codebook is in Appendix N.

Findings

The results of the interview and focus group analyses are presented below. These results describe the *components of district evaluation systems* (RQ1), *implementation differences across districts* (RQ2), reported *benefits* and *challenges* of the evaluation system (RQ3 and RQ4), and participants' *recommendations* to address challenges (RQ5). The qualitative findings are presented first in bulleted form, followed by a detailed narrative.

Frequency counts are *not* included in this section of the report. Given that interview and focus group sample sizes are much smaller than the survey sample, including frequency counts may diminish the importance of highlighting the perspectives shared here (i.e., the lived experiences of those interviewed individually or as part of a focus group). As Hannah and Lautsch (2011, p. 20) acknowledge, "By presenting numbers, qualitative researchers risk undermining the legitimacy of any insights they derived from small amounts of qualitative data." Instead of including frequency counts, terms such as "some educators who participated in focus groups" are used in instances where a few participants shared the perspective given. Terms such as "many/most educators" are used in reference to perspectives repeatedly shared by participants across relevant qualitative data sources.[20]

Key Findings

RQ1: Components of District Evaluation Systems

- Educators across multiple roles and districts reported that summative evaluation ratings were based on 70% professional practice and 30% student growth. Interview and focus group participants attributed the decision-making process behind the weighting to two factors:
    - collaboration between administrators and teachers at the state and local level
    - compliance with state requirements

[20] As a reminder, a variety of data sources are included in this report to offer different levels of detail. Examining them in relation to one another yields richer interpretations and more informed recommendations (Patton, 2002).

- Most teachers, administrators, and PERA joint committees described how their districts used Charlotte Danielson's *Framework for Teaching* (2007)[21] to assess professional practice. Fewer educators stated that their district used Marzano's framework (2013a).

RQ2: Implementation Differences Across Districts

- According to interview and focus group participants, student growth is assessed in a variety of ways across districts in Illinois, including Type I, Type II, and Type III assessments.[22]
    - Some teachers and administrators reported that their districts use an "all-in" approach to student growth, relying on the same standardized measure to determine the student growth rating for all educators in a school, regardless of the grade level or content taught.
    - Participants in "all-in" districts described a mix of benefits (e.g., all educators were rated using a common assessment) and drawbacks (e.g., some educators were rated based on student performance in a content area they did not directly teach) to this approach.
- Some districts have narrowed the focus of their observation rubrics, which has allowed evaluators to provide teachers with more specific and actionable feedback about their professional practice.
- Teachers and administrators expressed mixed views on the utility of preobservation and postobservation conferences.
- Some teachers, administrators, and joint committee members noted that their districts sought to improve the evaluation system in response to teacher and administrator feedback. Most notably, these districts tailored observation rubrics to job roles and used calibration protocols to improve reliability between evaluators.

[21] Charlotte Danielson's *Framework for Teaching* has a number of iterations, beginning with the earliest version in 1996. The 2007 version is cited in this report with the recognition that districts in Illinois have incorporated more recent versions as part of their evaluation systems (e.g., 2013; 2022). The 2007 version is cited unless an updated version is explicitly referenced by a study participant.

[22] The Illinois Administrative Code (see part 50:30 of Illinois General Assembly, 2022-a) defines a Type I assessment as "a reliable assessment that measures a certain group or subset of students in the same manner with the same potential assessment items, is scored by a non-district entity, and is administered either statewide or beyond Illinois." A Type II assessment is a district-adopted or district-developed assessment used by all teachers in a given grade or content area. A Type III assessment is "any assessment that is rigorous, that is aligned to the course curriculum, and that the qualified evaluator and teacher determine measures student learning in that course."

RQ3: Benefits of Evaluation System

- Teachers and administrators appreciated that observation rubrics, such as the Danielson Framework for Teaching, are research-based. Educators noted that this framework provides a strong vision for high-functioning classrooms.
- Teachers found benefits in the evaluation process when it included coaching, professional development opportunities, and specific and actionable feedback.
- Teachers noted that feedback tends to be more specific and actionable when their evaluator focuses on a select few components or domains of the Danielson framework.
- Some teachers preferred using locally developed assessments (e.g., Type II or Type III) over standardized assessments to assess student growth because these assessments can be tailored to their specific students and the content they teach.

RQ4: Challenges With Evaluation System

- Administrators and teachers expressed concerns that the evaluation process is burdensome, cumbersome, and time consuming.
- Administrators and teachers identified scheduling challenges associated with the evaluation process, as well as challenges related to evaluation timelines (e.g., evaluators find it difficult to complete a large caseload of teachers' observations by March; teachers find it difficult to demonstrate all indicators on the observation rubric within a 30-minute class period, etc.).
- Despite efforts to standardize the evaluation process, teachers and administrators described challenges related to observer subjectivity, bias, and inconsistency.
- Teachers reported increased stress and anxiety related to the evaluation process.
    - In some cases, teachers also noted that the evaluation process created adversarial relationships between evaluators and teachers.
- Administrators and teachers reported challenges with student growth measures, especially SLOs (e.g., time-intensive process, easily manipulated, etc.).
- The COVID-19 pandemic disrupted the cadence of the evaluation cycle for some tenured teachers. Some teachers felt that the pandemic also negatively impacted their overall evaluation ratings.
- Some joint committee members reported a lack of flexibility when implementing PERA to comply with state policy and regulations.

RQ5: Stakeholder Recommendations

- To address the time and workload challenges, teachers and administrators suggest a school- or districtwide effort to target a few specific domains of the observation rubric.
- To reduce administrative loads, some educators recommend using online platforms such as EvaluWise to gather and store evaluation data and assist with the evaluation process overall.
- To address interrater reliability issues and bias, some teachers recommend assigning multiple evaluators to work with each teacher. Other teachers recommend assigning evaluators with content and/or grade-level expertise to evaluate teachers in similar areas.
- Many administrators and teachers recommend instituting more regular, less high-stakes observations of professional practice to make evaluations more humanizing, authentic, and relational.
- Some administrators and teachers recommend that districts use instructional coaches to observe teachers and provide ongoing feedback, but keep instructional coaching separate from the evaluation process (i.e., coaches should not be evaluators).
- Some early contributors, administrators, and teachers recommend that the evaluation process incorporate observations and measures of student growth across several years.
- If changes to state policy are made, early contributors suggest that districts that are higher performing or better resourced be tapped to implement first.

RQ1: What are the components of districts' teacher and administrator evaluation systems?

To address RQ1, interview and focus group transcripts were analyzed for participants' descriptions of their district's evaluation systems, as well as insights on the decisions that informed the development of these systems. Findings associated with RQ1 mirror several of the survey findings and therefore help to explain and contextualize evaluation components commonly used across Illinois. In some cases, the findings associated with RQ1 also offer more detail regarding dissenting opinions, helping to explain why some educators' perspectives differ from the majority's regarding their district's evaluation components.

Common Elements Within District Evaluation Systems

Corroborating survey data, most educators, including teachers, administrators, and joint committee members, described the typical weights of measures in their district's evaluation plan as 70% professional practice and 30% student growth, with a mix of Type I, II, and III assessments used to measure student growth. Most participants attributed their district's rationale for the weighting of evaluation components to two factors: (a) collaboration between

administrators and teachers, often via the PERA joint committee; and (b) compliance with state requirements. While teachers and administrators reported several challenges associated with the different components of their district's evaluation system (described in greater detail in RQ4 below), joint committee members offered limited details as to how their district worked to address these challenges.[23]

During focus groups, teachers and administrators most often described using Charlotte Danielson's *Framework for Teaching* (2007) to inform observations of professional practice, with a few educators using Marzano's framework (2013a). While none described using other frameworks, some participants indicated that their districts had adapted Danielson to streamline observations and to focus on specific elements of professional practice rather than the entire framework. Other differences in how districts implemented components of the evaluation system are described in greater detail in RQ2.

Mixed Perspectives on Evaluation Components

Teachers and administrators offered mixed perspectives about the validity and fairness of their district's use of observational rubrics to assess their professional practice. Participants with more positive views agreed that the frameworks allowed for comprehensive observations and feedback, as well as a vision for effective instruction. For example, a joint committee member stated, "We all value . . . those conversations we have in a postobservation conference where we're talking about teaching and how to make some adjustments and improve and make recommendations. That's what I enjoy most from the evaluation process . . . it's about moving forward and getting better."

However, those teachers and administrators who offered criticisms of the evaluation system suggested that the observation process was unfair, that the components and descriptors attached to observation rubrics were insufficient, and/or the rubrics were overly broad. One teacher claimed, "What one evaluator thinks is an 'Excellent' could be what another evaluator thinks is a 'Proficient,' even though there's descriptions [in the rubrics]." Another explained that the evaluation process takes their focus away from teaching: "The evaluation system, it takes a lot of time. It takes a lot of time away from the classroom. It is burdensome." Many educators shared that both observations of professional practice and student growth factors were susceptible to observer bias or idiosyncrasies, an issue explored in greater detail through RQ4.

[23] Please note that the review of evaluation plan did not shed light on how the joint committees worked to address challenges; rather, the evaluation plans provided insights about the evaluation systems and processes that were in place at the time of the review foreach district.

Factors Informing Weighting Decisions

Early contributors whom we interviewed (including teachers and school/district leaders, as well as members of professional organizations such as the Illinois Federation of Teachers and Illinois Education Association) and school administrators who participated in focus groups agreed that the negotiation process between district leadership and teachers played a key role in determining the 70/30 weighting system that is prevalent across Illinois districts. Negotiations at the local level mirrored reports from early contributors we interviewed, who described the minimum 30% student growth requirement as a product of negotiations between multiple groups (e.g., union representatives, district leaders, and attorneys involved in writing the legislation) at the state level. Findings indicate a divide at both the state and district levels between those who valued the inclusion of student growth measures in the evaluation process and those who felt student growth was an inappropriate measure.

At the state level, several early contributors described the process of negotiating the 30% student growth requirement as a fair balance, given the different perspectives on teacher evaluation at the time of PERA's drafting. For example, one early contributor involved in PERA's inception and implementation described the process as follows:

> [W]hat we wanted to do is push districts into a 70/30 split, and that was all negotiated. It was like sausage being made. That would have significant impact, but also still have the practice be the main component. So, 70% of most evaluations are based on professional practice . . . Most of them are formed off the Danielson framework. And then the last, the 30% [student growth] . . . if you didn't agree to something like that, you ended up at a 50/50 split. So, that was our way of forcing districts to get to the table . . . because the unions would come in and the management would come in, and nobody wanted 50/50.

District-level participants—including administrators and teachers—described a similar negotiation process between district leadership and teachers. As one teacher explained, "[E]very district makes a different decision with their union . . . and for some districts, if you have great leadership and you have great union leadership and you guys come to an agreement, wonderful, right?" An early contributor added that from "teachers' unions . . . there's been a huge pushback on assessment measures."

Indeed, participants' descriptions of the relationship between district leadership and teachers varied. In some cases, the relationship was trusting and productive; in others, the relationship was strained and even adversarial. Participants—including teachers, administrators, and early contributors—explained that teacher representatives wanted a lower emphasis on student growth in the summative evaluation rating or a higher emphasis on supporting teachers who

needed help. As one teachers' union member explained, early on in PERA implementation, "One issue that arose was what percentage of the rating should be based on test scores . . . the teachers' unions, and other like-minded people . . . didn't want that to be any part of the equation because there's so much evidence that test scores are tied to socioeconomic factors." Though districts reported complying with the 30% minimum weight for student growth, the actual calculations and decisions about which assessments to include were common points of compromise and, at times, disagreement. Analyses of interview/focus group data suggest that the relationship between district leaders and teachers strongly influenced how districts decided to assess and weight the student growth component of the evaluation.

Emphasis on Compliance With the Legislation

In addition to collaboration between teachers and administrators at the state and district levels, multiple joint committee members, early contributors, and teachers described compliance as an important factor in the makeup and implementation of the evaluation system at the local level. One joint committee unanimously agreed that compliance was the underlying factor in how their district developed their plan, with one administrator explaining that compliance was their district's "original goal," even as they also took the view of "evaluation as a way of growing, not as a way of got you."

Both administrators and teachers shared that prioritizing compliance often seemed at odds with efforts to improve professional practice. One teacher characterized the balance between "the things that we're doing just for compliance" versus "things that we're doing to really help us improve." An early contributor suggested that since the passage of PERA, the evaluation process was "becoming more compliance-driven, less growth-driven" across the state. Other participants noted "loopholes" in the evaluation system that enabled educators to "get a proficient evaluation without that growth mindset." One early contributor explained districts' emphasis on compliance in this way:

> I do think when you pass bills, you pass laws, there's a message that's typically sent behind it and what's intended. I think districts felt like they had to build these really big systems that have required more time, energy, and money than maybe was necessary for what we're getting out of it. So what I would tell you, coming back to what you were asking about there with regards to effectiveness, I do feel like we are not leveraging this type of process to really focus on the real growth of the individuals, that it has become more about the compliance of meeting the statutory requirements because there's a lot at stake here.

Challenges related to prioritizing compliance are examined through RQ4 below.

RQ2. How do districts differ in their implementation of these components, and what factors (e.g., district needs, characteristics of the district) do stakeholders identify as potential reasons for these differences?

To address RQ2, interview and focus group transcripts were analyzed to identify differences that participants reported regarding how their district implemented components of the evaluation system. Analyses revealed that districts adopted varied approaches to the implementation of student growth measures, observation rubrics, and pre- and postconferences. Some joint committees reported that their districts developed customized observation rubrics for specialized roles, provided training to calibrate evaluators, and narrowed observation rubrics to provide targeted and actionable feedback. Although several districts emphasized compliance with the legislation (as noted in RQ1), others adapted components of their evaluation systems in response to educator feedback.

Student Growth Variations

As identified in survey data as well as the review of district evaluation plans, student growth is assessed in a variety of ways across districts in Illinois. While some districts incorporate locally developed assessments, others rely in part or completely on norm- and/or criterion-referenced assessments, such as NWEA MAP (Type II) or the Scholastic Aptitude Test (SAT; Type I). Educators, including joint committee members, teachers, and administrators, expressed varied perspectives on this component of the evaluation system. Some appreciated the flexibility afforded to them when developing SLOs, while others noted that the SLO process was time consuming and not worth the effort. Multiple educators expressed concerns about the fairness of using standardized assessments to rate a teacher's effectiveness.

A noteworthy example of district variation regarding student growth is the use of the "all-in" approach. Participants described this approach as using the same standardized measure to determine the student growth rating for all educators in the building, regardless of grade level or content taught. Some educators appreciated this approach. For example, one joint committee member observed, "We've changed [how we assess student growth] over the years, and we let teachers have a say in what they want it to be, and we do the all-in approach." Others expressed concern about an "all-in" model. As one teacher explained, "The overall performance of everyone, preK to 12, is factored into the score of every single evaluation done for the year," and that this was "an uncomfortable topic" that administrators "just skimmed over."

Regarding the selection of assessments, the amount of choice that teachers had varied across districts. One teacher with experience working in two different districts in Illinois shared the following:

> [In] the previous district I worked in, they negotiated through their contract [to] use the MAP testing. If you were a math teacher, if your math students showed growth on the MAP test, that was one of the benchmarks, and then you had another one you could choose. Whereas the district I'm in currently, we just have what the [joint] committee agrees on, [which are] these [unnamed] tests.

In comparison to districts that used standardized test scores exclusively to assess the student growth component of the evaluation, teachers working in districts that incorporated an SLO process indicated that these types of student growth measures allowed for some flexibility but were subject to manipulation. Because discussion of SLOs often focused on benefits and challenges, we detail this topic further in sections RQ3 and RQ4 below.

Variations in Observations of Professional Practice

Many participants across multiple roles agreed that the breadth of their district's evaluation system was both comprehensive and overwhelming. Some districts worked to address this issue by narrowing the focus of observation rubrics. Though not a common phenomenon, streamlining specific rubric components allowed evaluators to provide teachers with more targeted feedback. This approach also reduced the burden on administrators' time, thereby making evaluation caseloads more manageable. For example, one administrator shared the streamlined approach in their district:

> Instead of evaluating on the entire Danielson framework . . . we focus on seven components that are adopted by the district as a whole. Then each building has the option to choose one additional component that has been decided upon that they feel is a focus connected for their specific building in partnership with our school improvement plans.

Despite this example, analyses of interview and focus group data suggest that districts do not consistently adapt their evaluation systems to meet local needs. As one early contributor commented, "I do believe that districts probably haven't utilized some of the flexibility they do have in the statute."

Educators with experience in more than one Illinois district reported variation in districts' approaches to observations in general. These differences impacted the educator's overall evaluation experience. For example, one teacher from a suburban district described stark differences between the districts he had worked in: "[T]he district I'm in now . . . evaluations are there to help you improve and this is a collaborative effort, versus coming from a gotcha district where the union would stand outside your evaluation to make sure that everything was up to par, and everyone was always fighting." Another teacher in a city-based district explained

how their administrators exercised flexibility in utilizing evidence outside formal observations of professional practice to determine their overall rating:

> I came from a district before my current district where teachers and administrators would just walk in and out of your classroom. It was never a huge, big evaluation thing. So, the anxiety level was low in terms of that. I really think that our administrator took the time to kind of know what was going on in the classroom outside of the evaluation tool. So even if it wasn't something he specifically witnessed or saw in that evaluation day or through walk-throughs, he still had evidence that he could pull from as far as whether it be from artifacts or his personal observations.

In this case, an evaluator's willingness to incorporate evidence beyond a formal observation reduced the high-stakes nature of the evaluation process. The use of contextual evidence in an observation was a point of scrutiny amongst teachers and evaluators, as some described such information as vital, while others pointed out that their rubrics required a small window of observation with no other evidence factored in.

Adapting Observations of Professional Practice to Job Roles

In response to feedback from teachers and administrators, some PERA joint committees developed observation rubrics tailored to specialized job roles (e.g., special education teachers, school counselors, literacy interventionists). These changes accounted for the fact that the daily practice of some educators, such as an early childhood teacher working with students with developmental or cognitive delays, looks quite different from their colleagues' daily practice.

One special education teacher described how the use of adapted rubrics enables teachers to better explain their unique job responsibilities to their evaluator:

> [What] I have appreciated through the whole process is [the fact that the evaluator] is learning as he goes along, and his feedback and his conversations with me . . . have changed and grown and become more helpful for me, I guess. I don't know if helpful is the word I really want, but he's learning more about how special education fits into the evaluation process and how he has to look at it a little bit differently than I would think a traditional or general education classroom was. So I've appreciated the effort he's put into learning and growing through this process with us.

While some teachers indicated that the observation rubrics their evaluators used were not appropriate for or relevant to their job role, most participants, including teachers and evaluators, agreed the content of the rubrics was appropriate. Nonetheless, participants acknowledged that because observation rubrics are comprehensive, using them as part of the

evaluation process can be overwhelming and burdensome. This challenge is explored in greater detail through RQ4.

Variations in Pre- and Postobservation Conferences

In some districts, preobservation and postobservation conferences are important tools to encourage dialogue between teachers and evaluators, as well as to provide teachers with feedback to improve their instruction. In ideal cases, these conferences provide dedicated time for evaluators and teachers to reflect on evidence, identify strengths, and determine areas of growth. However, for some educators, the timing of observations and pre-/postobservation conferences was limiting. For example, if a conference occurred during a teacher's planning period and the evaluator was addressing another issue in the building, the opportunity to discuss insights was cut short. The challenges posed by scheduling and timing are explored in greater detail through RQ4.

Educators also expressed concern that the formal observation process rarely captured a complete picture of a teacher's professional practice, as one administrator explained:

> If you're just doing your district's process, there's so much more that goes into it . . . because they [teachers] are going to show you all the song and dance when it's a scheduled observation . . . I'm in the classroom all the time . . . if you are just doing the minimum evaluation process, I don't think it's fair or accurate. It's all the other extra time and effort that you put into it that makes it fairer and accurate.

In general, the utility of pre- and postobservation conferences seemed to be linked to a participant's overall experience with the evaluation process. That is, for participants who viewed evaluations as an opportunity to help teachers grow and improve, pre- and postobservation conferences were meaningful parts of the process. Conversely, for participants who felt that the evaluation process was a "gotcha" or an exercise in compliance, pre- and postobservation conferences were another hoop to jump through rather than a useful opportunity to reflect and grow.

Variation in Training

While most district- and school-based leaders did not mention interrater reliability, some reported that their districts used calibration exercises, or opportunities for evaluators to practice using observation rubrics alongside one another to become more consistent. In these cases, calibration is used as a learning tool for new and veteran evaluators. This process helps administrators develop a better understanding of the observation rubric and may help to increase the tool's utility. One early contributor had experience with this:

> I know my administrators here in the building have gone through at least two recalibrations. Well, they recalibrate every couple of years where they'll sit down together, and they'll watch some videos, or they'll have one or more evaluators come into a teacher's classroom and then they will go through the rubric together to make sure that they're kind of assessing things in the same manner.

While some administrators expressed concerns that calibration efforts served to highlight discrepancies in evaluation without truly addressing ways to mitigate it, others stated that their districts were making concerted efforts to ensure some degree of consistency between different evaluators. Issues of consistency are examined in greater detail through RQ4.

Variation in District Efforts to Improve Implementation

Despite the challenges that multiple participants attributed to different components of the evaluation system, there is mixed evidence whether districts have attempted to address common issues with PERA. One early contributor who regularly corresponds with educators in multiple districts described multiple changes:

> Across the state, we've seen a lot of shift[s] in the way we do evaluation because of PERA, and I think that's both supportive for administrators and teachers and students . . . I think we've done a lot of training around how to have crucial conversations, how to support educators, [and] how to support educators when supporting their students. And again, using a rubric or the Danielson rubric or Marzano, whichever rubric you're using, just to have that balance and that guidance for teachers, as they're thinking about improving their practice.

Despite this, focus group participants provided little evidence that most districts have adapted evaluation components. Two joint committee members claimed the following:

> **[Member 1]** I think we only meet to review this when they tell us. Last time, it was, we had to meet before September 30th, 2021, wasn't it? Or 2022. I mean literally that was a mandate. So that's the last time we met to review it. There's not much point meeting to review it if you really don't have the flexibility to review it and change it.
>
> **[Member 2]** Yeah, it's basically we do by the law what they tell us to do, and that's it.

These joint committee members were district administrators who claimed that they had little to no ability to change the evaluation system. As a result, the joint committee made few alterations to their district plan unless the state required them to do so. Another early contributor explained that, as a state, "We are not leveraging this type of process [evaluations]

to really focus on the real growth of the individuals, that it has become more about the compliance of meeting the statutory requirements because there's a lot at stake here."

Despite mixed reports, some districts' joint committee members demonstrated that it was possible to develop evaluation systems that were tailored to local circumstances, responsive to teachers' needs, and encouraged reflection and growth.

RQ3: What benefits have stakeholders experienced implementing components of the teacher and administrator evaluation system with respect to the following: obtaining useful feedback, supporting improvements to professional practice, informing policy decisions, and adapting to the COVID-19 pandemic?

To address RQ3, the research team analyzed focus group transcripts to identify the benefits that teachers and school-based administrators experienced through their district's evaluation process. It is important to note that consistent patterns of benefits across districts were not identified through the data analysis. Rather, the benefits cited by participants were situational, often linked to their district context or a specific evaluator, rather than tied to the evaluation system overall. Primarily, participants shared that their district's observation rubrics provided some degree of objectivity in the evaluation process, as well as a clear and consistent view of what high-quality instruction should look like. When districts offered coaching and professional learning options, teachers appreciated opportunities that supported their instructional growth and development.

Fair and Objective Measures

Some teachers and administrators agreed that their district's evaluation system was fair, adequate, and/or objective, specifically citing the use of observation rubrics and standardized student test scores as reasons. As one teacher in a town/rural focus group expressed it:

> I do feel like our district is fair and accurate in how they're evaluating all teachers to try to keep it as standardized, if you will, and not like, 'Oh, I just like that person better' kind of thing. They have hard evidence that they're using to make their decisions.

This teacher believed that evaluators' use of a standardized observation rubric, as well as the use of "hard evidence" to determine a teacher's performance rating, was a fair and unbiased approach. Not all teachers framed this approach as fair or beneficial, however.

Clear Rubric and Structures to Improve Instruction

Some teachers and administrators said the value of a clear observation rubric was that it provided a vision for effective instruction and informed coaching opportunities. When describing the Danielson Framework for Teaching, participants appreciated that it was

research-based and supported evaluators in identifying evidence of high-quality instruction. One teacher who served as an instructional coach offered this example:

> I think when I was using the Danielson rubric, it helped to actually improve my program, the instructional coach program . . . It gave some clear guidelines and then the conversation between my principal and I to help clarify my role. The rubric was very, very helpful.

In some cases, the evaluation process also provided a structure for coaching and feedback opportunities that teachers found valuable, as one teacher noted:

> [I]n the district that I'm in now, the biggest thing that they provide are those instructional coaches. The ability to just meet with the instructional coach, whether it be on something that you're being reviewed on or something else in relation to your class.

This evidence indicates that both coaches and teachers appreciated the structured approach of observational rubrics and evaluation processes, finding it beneficial for delivering and receiving feedback that supports professional growth.

Flexibility and Autonomy

Although not a prevalent trend, some teachers and administrators provided examples of how educators could leverage flexibility and autonomy when participating in the evaluation system. These examples were often described as a strength of the evaluation process. For example, one school-level administrator described some of the choices that teachers had in their district, and how these choices supported self-reflection:

> [A]nother strength, there's a self-evaluation tool on our program that we have . . . If the teacher completes it [the rubric] before we have our postevaluation meeting, then I see what they've picked for themselves . . . and then I pick mine . . . It's a good way to have a discussion about that . . . [and] . . . where they see themselves and why . . . [T]hat's one of the things I really do like.

Although most teachers in this case did not elect to complete the self-evaluation, this evidence suggests that the evaluation system can be leveraged to promote reflection.

How the SLO process was incorporated into the district evaluation system also gave teachers various ways to exercise their autonomy. In some districts, teachers shared that SLOs allowed them to better tailor the student growth component of the evaluation system to their grade and content area. Some also stated that SLOs enabled them to better understand how to improve as an educator.

In terms of flexibility, teachers appreciated when their evaluators incorporated evidence of professional practice beyond formal observations of professional practice. As one teacher expressed:

> I really think that our administrator took the time to kind of know what was going on in the classroom outside of the evaluation tool. So even if it wasn't something he specifically witnessed or saw in that evaluation day or through walk-throughs, he still had evidence that he could pull from . . . whether it be from artifacts or his personal observations.

Teachers and administrators both indicated that, to provide more relevant feedback, evaluators had to make independent decisions during the evaluation process. In a focus group, two school leaders extolled several aspects of their observation practices. When asked whether these benefits were part of their evaluation system, they explained that "we have to do extra" and that "it's a very tedious task." One explained further that "once you're done with your postconference, let's say you're done with your time . . . because this process is so tedious, I can't do the little follow-ups that would really guide growth." To provide more targeted observation feedback, these leaders recognized that they made individual choices, which they further explained were sometimes at odds with the requirements of the evaluation system.

Largely, the benefits experienced by teachers and administrators depended on district context as well as relationships between the evaluator and those they evaluated. Moreover, in districts where the focus of observations was more targeted, teachers stated that the feedback they received from their evaluator was more specific and actionable. In districts with strong, trusting relationships between teachers and administrators, the evaluation process is generally reported to be less contentious and more beneficial to teachers.

RQ4: What challenges have stakeholders experienced while implementing components of the teacher and administrator evaluation system with respect to the following: obtaining useful feedback, supporting improvements to professional practice, informing policy decisions, reported biases or inequities in system components, and adapting to the COVID-19 pandemic)?

To address RQ4, the research team analyzed focus group transcripts to help identify the challenges that joint committee members, teachers, and school-based administrators reported experiencing while implementing components of the evaluation system in their districts. In addition, interview transcripts from early contributors were analyzed to identify challenges that this group anticipated before districts implemented PERA and that they observed after districts implemented PERA. Across these varied participants, respondents identified several overarching challenges: concerns with subjectivity, bias, and inconsistency in the evaluation

process; barriers to making the evaluation process a meaningful opportunity to improve professional practice and student learning; and educators' experiences of burden, stress, anxiety, and distrust as a result of participating in the evaluation process.

Subjectivity and Bias

While some participants stated that they believed the observation rubrics supported fair and accurate observations of professional practice, many others stated that observations were subjective and biased, despite the use of rubrics. Some educators expressed that even when evaluations are based on clear descriptors and require evidence to be collected or documented, the observation process is inherently subjective because every evaluator brings their own interpretations and opinions to the table. One teacher from a low-income district in a city experienced this subjectivity this way:

> Even though it's all listed the same there on the document, it's like you can get the four people probably could watch you at the same time . . . They'd all have their own opinion about the same thing and they could all be bad or they could all be good, or two could be bad, two could be good of the same thing. I don't know. Just whoever's interpreting the document in their own way has their own opinion about it.

Another teacher from a high-income district in a city stated that their principal "had preconceived thoughts" when it came to evaluating teachers. This teacher went on to state that they were unsure "how much evidence is enough evidence to prove yourself" and that although they provided two to three pieces of evidence per domain on the Danielson rubric, their evaluator said that it was not enough evidence to be rated "Excellent." This teacher was left wondering, "How much is enough to say, I am checking these boxes?" Other teachers also described instances of an evaluator approaching the observation process with a predetermined rating in mind. Additionally, other teachers reported not knowing how much evidence was "enough" to either justify a performance rating or to move to a higher performance category.

In addition to perceptions of unfairness due to evaluators' differing interpretations of an observation rubric, teachers and administrators also linked bias and unfairness to the population of students that a teacher serves. Participants viewed this form of bias and unfairness in observations of professional practice and student growth measures. As one administrator stated:

> Teachers who have older, homogenous, highly engaged student populations are likely better able to achieve highest eval levels. Special education classrooms, primary classrooms, teachers with a "tough group" this year . . . a one-done eval is not accurate.

A number of teachers serving specialized populations (e.g., multilingual learners; students with individualized education plans) or specific grade levels and content areas reported that the evaluation process was less fair for them than for their colleagues. In these cases, perceived unfairness was often linked to an evaluator's inflexibility when interpreting a rubric, an evaluator's lack of experience or expertise in the area(s) being evaluated, or a teacher's inability to meet the expectations of a rubric given the students they serve and/or the grade level or content area they teach. Regarding unfairness in student growth measures, one teacher from a high-FRPL district in a city explained, "Standardized tests don't show if the student has grown in the areas that their IEPs [individualized education plans] are." This teacher felt that standardized assessments were unfair for her and her students since the purpose of those assessments (i.e., measuring proficiency on grade-level standards) was not aligned with the goals outlined in her students' individualized learning plans. Another teacher who served bilingual students noted that her evaluator did not speak the language she was observed teaching.

Inconsistencies Across Evaluators

Teachers and administrators noted that, in addition to reports of observer subjectivity and bias, inconsistencies across evaluators posed a significant challenge in their districts. Some linked this issue to inadequate training for new evaluators and a lack of calibration among evaluators working in the same building or district. One teacher identified an influx of new administrators as contributing to inconsistency:

> My district has had new administrators very frequently, and they're new to administration . . . [I]n the summer, [they] get taught how to do this [evaluation process]. Then they come to my building that has over 200 teachers in it, and they're expected to do this. And so the evaluator was telling me my first couple of years here, 'I'm just learning how to do this, so bear with me.' Would I say that's adequate? No, they're not even given enough time to learn how to do it.

Because of inadequate training and lack of calibration opportunities, teachers and administrators noted that evaluators within and across schools, grade levels, and content areas do not consistently interpret or use their district's evaluation rubrics. These inconsistencies contributed to teachers' perceptions that evaluations are subjective.

Student Growth Challenges

Administrators and some teachers reported challenges with student growth measures, especially the SLO process. Both parties acknowledged that the SLO process can be easily manipulated, as revealed by this administrator from a suburban district:

> SLO has nothing to do with teacher instruction. So many times, over the last dozen years it's been, 'I'm going to give them a test, they're going to take it. They're not going to do well. I'm going to teach them the stuff. I'm going to give them the test, they're going to take it, they're going to do well. I'm going to get my 30% excellent and that's going to I'm done. Check, check, check.' It's a task.

Teachers and administrators tended to agree that the process of writing an SLO was time consuming (a point explored further in the next section, Evaluation Burden). Furthermore, some teachers and administrators reported that SLOs rarely led to meaningful instructional improvements or improvements in student learning. As an administrator in another suburban district explained, "It seems like a lot of energy towards something that doesn't move the needle on growth for instructors or growth for the system." Joint committee members and early contributors to PERA also acknowledged challenges posed by student growth measures, and early contributors acknowledged that these challenges were anticipated at PERA's onset.

Evaluation Burden

Even in cases where benefits of evaluation were reported, teachers and administrators described the evaluation process as time intensive, burdensome, and cumbersome. Many teachers and administrators agreed that those who experienced the greatest burden were evaluators with large caseloads (20+ teachers), newer evaluators, and newer teachers. Participants noted that the amount of paperwork required to complete an evaluation cycle is overwhelming, particularly for those who are newer to their roles. While some districts have adopted an electronic system to streamline the evaluation process, participants in those districts reported mixed results. Although one joint committee explained that teachers and administrators were more satisfied with the online system, administrators and teachers in other districts reported that their systems remain time intensive.

Overwhelmingly, teachers and administrators recognized that the evaluation process felt like an exercise in compliance rather than a process that meaningfully improved teaching and learning. We frequently heard participants use terms like "checking boxes" and "jumping through hoops." For many, the focus of evaluation was adherence to state and district policy. It could feel a bit mechanical, as one teacher explained:

> I guess it kind of feels like a checklist. 'Yes, I'm doing this, yes, I'm doing this. Yes, I'm doing this.' I understand that there are, we hear it in the news, we hear it, we see it all over the place complaints that there's some people that aren't necessarily doing their jobs to the standard that we would expect or we would hope. But I feel like this never-ending prove-myself is because there's one bad apple in the bag. The rest of us kind of

> have to jump through all the hoops to make everybody, make the media, society, whatever, feel better.

Evaluators also admitted that "the intellectual load of writing an evaluation that empowers a teacher is heavy," and that it becomes increasingly heavier as your caseload increases. Administrators linked several related challenges to large evaluation caseloads. Perhaps most significantly, some evaluators indicated that by having too many people to evaluate, they do not have enough time for regular, informal observations of teachers' instruction. Evaluators expressed that being able to conduct more frequent, informal observations would provide a better sense of what typical instruction looks like and would contribute to more informative conversations with teachers on ways to improve their professional practice. Multiple teachers agreed that ongoing observations of their instruction would be preferable and more helpful than the formal observation process. Without these more regular opportunities for an evaluator to observe and provide feedback on a teacher's instruction, participants may be inclined to continue to view the evaluation process as a matter of compliance. That is the position one teacher from a suburban, high-FRPL district has taken:

> The most honest thing an evaluator told me about this process is that there is so much paperwork and hoop jumping on both sides that the focus becomes strictly on compliance with the law and not thoughtfully evaluating a teacher.

Although participants largely agreed that observation rubrics provided a comprehensive picture of effective professional practice, the benefits of using these rubrics did not outweigh the burden. Several participants stated that the time it took them to use the rubric to prepare for a pre- and postobservation was not worth the feedback they received; even among those who appreciated their feedback, none claimed that it was worth the effort. One teacher reported that the evaluation process feels more like they are being told whether they are doing their job or doing it well, rather than supporting them to improve. Other teachers noted that they are responsible for identifying their own plan for professional growth and development rather than receiving guidance and support in this area from their evaluators. This aligns with administrators' reports that the amount of paperwork involved in completing the evaluation process made it difficult to provide teachers with targeted, personalized feedback.

Participants also reported that the SLO process was time intensive. While SLOs could be helpful for a teacher to monitor students' growth over time, especially when standardized assessments were not available, some teachers stated that SLOs were not worth the time and effort involved. Overall, challenges related to the burden of evaluation may explain the propensity of teachers and administrators to view the process as a matter of "checking boxes" and "jumping

through hoops," rather than a process that meaningfully informs a teacher's professional growth and development.

Scheduling and Timing Challenges

Both administrators and teachers named scheduling evaluations and meeting evaluation timelines as challenges, as a joint committee member acknowledged:

> I think another challenge is just the sheer scheduling of all the evaluations . . . as a teacher, I need to know that this is the year that I'm not being evaluated, but I need one informal and then next year I'll be evaluated . . . I can't imagine it from the lens of an administrator who has, in my school, she has 30 plus teachers and then so many assistants on a different system, and I don't even know how that chart would even look on a piece of paper to keep organized about that. I can hardly remember what I need to be evaluated on in certain years because of the gaps of time in between. I think that's just tricky . . . We're getting a lot of new teachers in our district, so that's just adding sheer numbers of evaluations as well.

This quote reveals perceived challenges of both keeping track of evaluation requirements and finding time to schedule all the necessary components. Participants also noted that the timeline to complete an evaluation is difficult because most summative evaluation ratings are determined in March. This is especially challenging for teachers who are evaluated annually, as the student growth measure does not account for a full year of instruction.

Teachers also described challenges related to being able to "prove" their effectiveness within the time constraints of a formal evaluation, and evaluators discussed challenges that occur when other job-related duties interfere with their observation schedule (e.g., a student behavior issue, a parent phone call, etc.). This can result in feelings of confusion, as a teacher from a town/rural, low-FRPL district shared:

> I feel a little pressure to get it all in one 45-minute slot or less. I don't know if anyone else feels that way. Maybe I'm still feeling new to my role and it's like, 'Oh, my evaluation's over now, but did I even get it all in? Did I prove myself?' . . . I know there's restraints to what you can do, but also feeling like, okay, that was one day out of the semester. I get evaluated twice a year right now. Did I prove myself or was it just a show that day and I'm not doing it the rest of the days?

This teacher's experience of worrying about proving herself within one class period connects to other teachers' concerns about how much evidence is "enough" during a formal observation to receive a particular rating. This challenge, related to scheduling and timing as well as an

evaluator's interpretation of what counts as evidence, could be exacerbated when an evaluator is unable to observe the full class period due to extenuating circumstances.

Stress, Anxiety, and Distrust

Educators in general, and teachers in particular, named challenges related to several affective dimensions of evaluation, including the job-related stress, worry, and anxiety they experience. These issues can contribute to distrust between evaluators and teachers. One joint committee member acknowledged that stress and anxiety can occur for teachers and evaluators regardless of their experience level and prior performance ratings:

> Currently, time and stress levels are huge around these. We just had someone going through it today and she was nearly coming out of her skin and she's a fantastic educator. And just the circumstances that were going on in our building, it ended up being a different person evaluating her than she had first thought, which I mean, both of the people who were possibly going to be evaluating her would've thought that she was doing wonderfully . . . And I think that the administrators have a whole lot of work to do to complete the evaluations and have all of the necessary documentation, and then trying to line up the dates and times and when they can be done or not done and the differences that different schools, from what you hear, different schools do it and approach it differently even amongst the same district.

Teachers also revealed that their negative experiences with the evaluation process may help to explain educator shortages, as demonstrated in the following quote from a teacher in a suburban, high-FRPL district:

> I would just really come back to the whole thing of what this has really contributed to, the current climate that we're all facing right now. And you're having a hard time keeping people right now because of that, just, adversarial nature to this whole thing. And it's a lot of why I think a lot of people aren't getting into the profession right now.

As a teacher from a suburban, low-FRPL district shared, "Our job as educators is tough to begin with, but add the stress of feeling that the administrator evaluating you is out to 'get you' and that is why it is so tough to continue to retain teachers."

Some teachers associated the high-stakes nature of evaluations as an added source of stress and worry:

> Before this model [Danielson], I felt better about my evaluation process. It was that I was observed by my principal for a couple different lessons. It felt more relaxed. It felt like it wasn't so high stakes with my reputation, I felt better about it. And then this

> model is so stressful and it's so time consuming and it doesn't feel useful, and I thought I'd heard that originally it was created for teachers to be self-reflecting not for principals to judge.

The final sentence of the quote above captures another common sentiment shared across focus group and interview participants: the Danielson framework is not intended to be used as an evaluative tool. Notably, one administrator also shared that the high-stakes nature of evaluation was contributing to a lack of risk-taking and growth among his educators:

> Anytime you're going to put a rank or a score into something, basically you're putting people somewhat fearful of taking a risk . . . And risk is the way we grow. We don't grow till we take risk. But if you've got a system in place where people are fearful of that score they get at the end, or that's what they're worried about, they're fearful of taking the risk or they're trying to manipulate to get to the score, or they're not really taking the risk that leads to growth. So that's one of the things that I see [as a] problem with this.

Although some early contributors stated that PERA was designed to improve teachers' professional practice and promote their professional growth and development, according to interview and focus group respondents the evaluation system as designed may make it difficult for teachers to take risks or innovate in ways that could lead to more meaningful improvements.

COVID-Related Challenges

In the development of RQs and subsequent data collection, PEAC and ISBE anticipated that focus group participants might report COVID-related challenges tied to the evaluation process. However, many of the challenges that respondents named were not exclusively tied to the COVID-19 pandemic. Cases where direct ties were identified were presented in the general form of, "COVID-19 interfered with evaluation timelines and with student growth measures" (i.e., all teachers received "Proficient" ratings on the student growth component of the evaluation). For some teachers, this interference resulted in obtaining tenure with a less extensive evaluation process (e.g., they did not have to write and track an SLO). In other cases, COVID-19's disruption resulted in added stress and anxiety for teachers.

A small number of teachers shared that the universal "Proficient" designation for student growth during the pandemic resulted in their summative evaluation scores shifting. Specifically, prepandemic they may have been rated "Excellent" overall, but when the student growth component was paused, they received an overall rating of "Proficient" instead.

> I know my previous evaluations prior to COVID, I had scored a Level 4 [Excellent], and then we were not allowed to be evaluated. So although my ratings over the 2 COVID years were down to a Level 3, and not that it's a big deal, but I work hard for my Level 4s . . . knowing that I didn't even have an opportunity to present myself to possibly have those higher ratings, I feel is a detriment because now that's on my record.

Although some teachers suggested that rating changes such as this had a negative impact on them, others described the pauses and modifications to evaluations during the pandemic as helpful, as described above when teachers did not have to develop and monitor an SLO or when they obtained tenure through a less extensive evaluation process. These educators appreciated how their district handled evaluations during the pandemic, with one participant explaining, "[In] my district, this is the one thing that they did good in."

In addition, COVID-19-related interruptions interfered with the cadence of the evaluation cycle for tenured teachers, as one teacher from a ow-FRPL district in a city explained:

> Evaluations just held for a year . . . And so then when we came back in 2021–22 in person, full-time, those of that had just been summatively evaluated were summatively evaluated again. So I had two back-to-back summatives, and some people went 3 years without a summative. So it was weird because your rating from 3 years ago just held with you until your next one. And then some of us were evaluated back-to-back school years that we were in person.

Although COVID-19 disrupted the evaluation process in many districts, the challenges associated with this disruption were not cited as often as others. However, participants described the changing landscape of schools in the wake of the pandemic as a challenge, one that extended beyond the evaluation process (e.g., the behavioral and academic needs of students are more significant than they were prior to the pandemic).

Lack of Flexibility to Adapt the Evaluation Process

Participants described not only the challenges they encountered during PERA implementation but also the steps their districts took to address these challenges. Such steps were typically under the purview of the PERA joint committee. Some committee members felt empowered to adjust their evaluation plans over time, while others reported that they did not have much flexibility to adapt because they needed to maintain compliance with the law.

Although PERA allows local education agencies to develop evaluation plans that meet the requirements of the law but tailored to district priorities, some joint committee members who participated in focus groups reported they did not view this as a "real" option. To adhere to the

requirements of the law, some district joint committees found it difficult to develop evaluation plans that were responsive to district needs and priorities.

RQ5: How do stakeholders suggest addressing reported challenges? What supports, resources, and changes are recommended?

To address RQ5, focus group and interview transcripts were analyzed for identified recommendations that participants suggested in relation to challenges (e.g., resources, supports, and changes that participants recommended to improve the evaluation process). These findings are summarized below.

Streamline Evaluation Workload

To address workload and timeline challenges associated with evaluations, especially observations of professional practice, multiple teachers and administrators suggested a school- or districtwide effort to target a few specific elements of the observation rubric. A more targeted approach can lead to more actionable and supportive feedback and professional development, as one administrator stated:

> I found the most value saying, I see a couple of areas where with some targeted focus on developing their instructional practice in that subdomain or two or three, they'll really benefit the student or the classroom in a really concrete way. And so that doesn't mean the other things aren't important, it's about attention and bandwidth.

Participants also recommended using online platforms like EvaluWise for collecting and monitoring evaluation data (e.g., observation data, artifacts, professional growth plans, etc.). Online tools may create efficiencies for educators and make the process more streamlined, thereby reducing administrative loads. Participants also noted other advantages of using an online evaluation platform. For example, teachers reported that they appreciated being able to access evaluations from previous years to inform their professional goals.

Improve Consistency Between Evaluators and Across Teaching Levels/Disciplines

To address interrater reliability issues and reduce observer bias, some districts assign multiple evaluators to work with each teacher. Some teachers reported that being evaluated by different evaluators reduced the perception of “luck of the draw” when evaluators are assigned.

Some teachers also recommended assigning evaluators with content and/or grade-level expertise to evaluate teachers in similar areas. Teachers serving certain populations of students and/or content areas outside of math, ELA, science, and social studies (e.g., special education teachers, elective teachers, and bilingual teachers) were more likely to voice concerns about evaluators lacking an understanding of the students and/or the content they teach.

Shift Emphasis to Growth and Development Rather Than Ratings

To address the tension between receiving an evaluation rating, which can be stress-inducing for some teachers, and supporting a growth-based mindset, some districts provide instructional coaches who observe teachers and provide feedback, but who do not conduct evaluations. Teachers attribute the culture of lower stakes observations and feedback from coaches as important to reducing the stress they experience when they are formally observed as part of the evaluation process.

Some educators also recommended that the evaluation process incorporate observations and feedback, as well as measures of student growth, across several years. They believe that this would support a more comprehensive view of teachers' practice and more meaningfully inform the feedback that teachers receive.

Prioritize Walk-Throughs Over Student Growth

Many administrators and teachers recommended instituting more regular observations of professional practice to make evaluations more humanizing, authentic, and relational and less high stakes. Some also suggested removing the student growth component altogether, so that evaluators can focus exclusively on observations and feedback. One administrator from a suburban district described the shift that would result:

> What would I trade if I could trade away the student growth component to be able in those first 2 years of a nontenured teacher to formalize at the state level or at a district level, there'll be three to five walk-throughs . . . it would be really powerful to be able to sort of say, I need to get a sense of the tenor of the classroom. I need to come in a number of times. And that would be different for someone who's tenured with a lot of experience . . . But yeah, I think it'd be incredibly valuable to spend more time formally, sort of seeing somebody early on in their career and less time collecting data tables from self-generated student growth quizzes. Walk-throughs over student growth.

As this example demonstrates, administrators value opportunities to provide teachers with ongoing feedback. They recommend changes to the evaluation process that would create more opportunities to provide meaningful feedback to teachers on a regular basis. To do this, districts may need to reduce other aspects of the evaluation process so as to not add more to teachers' and administrators' workloads.

Reprioritize Order of Implementation

If changes to the state policy are made, early contributors suggest tapping the districts that are higher performing or better resourced to implement first, rather than requiring the lowest performing districts to implement first. When PERA was passed, the lowest performing districts

in Illinois were required to implement first (Illinois General Assembly, 2024-a), and some early contributors believe this presented challenges for the most underresourced districts in the state. As one early contributor explained:

> There was such a push to get the 20% lowest performing school districts or schools to implement first . . . A lot of times, they were the lowest capacity. Their funding wasn't equitable. You take your lowest capacity districts, and you make them do some kind of complex reform first, that's really hard. And I understood the urgency people were pushing that for because, 'Oh, we got to get them. They're the ones who have the worst teachers because they have the lowest scores.' It's counterintuitive to do it that way, but then you have to think both the unanticipated challenge for that, and just an overall unanticipated consequence and challenge was that they struggled mightily.

Given these challenges, the early contributor quoted here suggested that it may be more effective to first test modifications to the evaluation system in districts with the capacity to implement those changes effectively. If those updates are successfully implemented in better resourced districts, they could then be rolled out to districts statewide.

Triangulation

In this section, the triangulation approach used to examine and synthesize evidence from multiple data sources, including the survey, interviews, focus groups, evaluation ratings, and evaluation plans, is described. By comparing these different data sets and identifying areas of convergence and divergence, a more nuanced understanding of educator experiences with their district's performance evaluation systems emerged. This approach allowed the research team to document and support the key findings presented below.

Methods

The triangulation approach involved comparing and contrasting findings across multiple sources of data to identify patterns, including commonalities and inconsistencies, to lead to more reliable and robust conclusions (Greene, 2008; Tashakkori & Creswell, 2007). To guide this process, the research team adopted an integrated mixed-methods design, in which both quantitative and qualitative data—in this case, data from the survey, interviews, focus groups, evaluation ratings, and evaluation plans—were interpreted alongside one another (Yin, 2006). Specifically, the research team conducted separate analyses of each data type, and then examined how the themes that emerged from each separate analysis supported and contradicted one another.

This method of analysis assisted in (a) determining the extent to which findings are consistent across multiple data sources, enhancing the overall quality and credibility of the results; and (b) identifying discrepancies between data sources, offering insights into the complexities and varied perspectives within the data. Triangulating multiple data sources results in a nuanced understanding of the opportunities and challenges facing educators as they engage with their district's performance evaluation system.

Findings

This section presents key findings (in **bold**) and supporting evidence for each research question.

RQ1. What are the components of districts' teacher and administrator evaluation systems?

Most district evaluation plans included two components: a measure of teacher professional practice and a measure of student growth. These two components are typically weighted at 70% and 30%, respectively.

Fifty-one district evaluation plans were reviewed and all included weights of 30% for student growth and 70% for professional practice. Many of the teachers, administrators, and joint

committee members who participated in interviews or focus groups confirmed that their districts adopted these weights.

Most districts measured the professional practice component using the Danielson Framework for Teaching or Marzano Teacher Evaluation Model, and 88% of the district evaluation plans we reviewed measured the student growth component using some combination of state or national standardized assessments and district- or locally developed assessments.

Many survey respondents (76.5%) reported that their district's evaluation system used observation rubrics to measure professional practice, and 8.4% reported other measures of educator practice. Interview and focus group participants reported that districts typically used the Danielson or, to a lesser extent, Marzano frameworks to measure professional practice. The reviewed plans used a mix of state or national standardized assessments and district- or locally developed assessments to measure student growth. Approximately 54.0% of the survey respondents reported that their evaluation system used districtwide and locally developed assessments to measure student growth, 28.8% reported standardized national and/or statewide assessments to measure student growth.

All districts identified professional development or remediation for educators rated "Needs Improvement" or "Unsatisfactory," but district evaluation plans did not typically articulate a process by which teachers rated "Proficient" or "Excellent" could access personalized coaching and/or professional development.

Most (70.6%) of the reviewed evaluation plans adhered to legislative requirements by prescribing professional development for educators rated "Needs Improvement" and remediation for those rated as "Unsatisfactory." Fewer plans (31.4%) described how evaluations would inform the professional development of educators rated "Proficient" or "Excellent."

Most (97.2%) educators who received ratings in 2022–23 were rated "Proficient" or "Excellent," with far fewer (2.8%) rated "Needs Improvement" or "Unsatisfactory." Because most educators received "Proficient" or "Excellent" ratings, they may not have received targeted development opportunities that could further advance their skills and effectiveness. As one educator shared, "Evaluations always seem to be an isolated thing, so not connected to larger professional development or coaching that's happening." Although teachers who participated in focus groups generally agreed that they could access professional development opportunities connected to their goals, many described needing to take the initiative themselves rather than rely on direct guidance from their evaluator.

RQ2. How do districts differ in their implementation of these components, and what factors (e.g., district needs, characteristics of the district) do stakeholders identify as potential reasons for these differences?

Almost all educators who participated in their district's evaluation system in 2022–23 stated that they were observed at least once as part of that process.

Most survey participants (97.6%) reported being observed at least "Once" during the previous school year. However, observation frequencies varied considerably with respect to tenure status and race/ethnicity. As one might expect, a higher percentage of nontenured teachers (97.5%) were observed "Two or More Times" than tenured teachers (67.4%). Additionally, the rates at which Black or African American (90.0%) and Hispanic or Latinx (90.9%) educators reported being observed "Two or More Times" were higher than the rate reported by White educators (77.4%).

Some districts have tailored their observation rubrics to meet the needs of educators in different roles or to allow evaluators and teachers to focus on targeted areas of professional practice.

During focus group discussions, observation rubrics such as Danielson and Marzano were deemed useful by both new and experienced teachers for understanding professional practice expectations. Educators in districts that tailored their observation rubrics to focus on targeted areas of professional practice reported receiving high-quality feedback. For example, one administrator shared, "Instead of evaluating on the entire Danielson framework or through the whole framework, we focus on seven components that are adopted by the district as a whole [and] each building has the option to choose one additional component . . . that they feel is like a focus connected for their specific building in partnership."

Some districts made other adaptations in efforts to improve the effectiveness and relevance of observational practices under the PERA framework. For example, some districts customized observation protocols for specialized roles, including special education teachers, school counselors, and literacy interventionists. Other districts emphasized interrater reliability by implementing calibration protocols and trainings to further ensure consistency in teacher observations.

Educators reported that pre- and postconferences were important tools. However, coaching opportunities were limited, and tenured teachers typically received less coaching than their nontenured counterparts.

In interviews and focus groups, some teachers, administrators, and joint committee members characterized preobservation and postobservation conferences as important tools for

supportive conversations between educators and evaluators. However, the prevalence of these conferences appeared to vary considerably by tenure status. Nontenured teachers were considerably more likely to participate in "Two or More" pre- (86.4%) and postconferences (86.1%) compared to their tenured peers (25.2% for preconferences and 29.4% for postconferences).

Coaching was limited, with only 30% of educators reporting that they received coaching. As one might expect, tenured teachers received less coaching compared to their nontenured counterparts. Survey results found a higher proportion of tenured teachers were "Never" offered coaching (79.9%) compared to nontenured teachers (56.6%). Conversely, nontenured teachers reported receiving coaching "Two or More Times" more frequently (29.3%) than tenured teachers (8.5%). Teacher focus group participants from low-FRPL districts provided more examples of coaching opportunities than participants from high-FRPL districts, and when teachers had access to coaching, they described the benefits. For example, as one teacher from a low-FRPL district commented, "The ability to just meet with the instructional coach, whether it be on something that you're being reviewed on or something else in relation to your class . . . has been a huge help."

During focus group discussions, teachers expressed appreciation for coaching opportunities when available and indicated they would value increased access to instructional coaching in districts where it is limited or unavailable. One teacher described how their suburban, low-FRPL district provided new teachers with access to a mentor who also offered coaching support: "Our new teachers each get a mentor who meets with them every week. They go through the evaluation process, give tips, answer questions, are a sounding board, coach, et cetera." Another teacher, also from a suburban, low-FRPL district, stated that they wished their evaluators would adopt more of a coaching perspective during the evaluation process. As she explained, "I think if our evaluators had more of a coaching mentality—let me come alongside you and let's do this together for the betterment of the kids—I think that would be different."

RQ3. What benefits have stakeholders experienced implementing components of the teacher and administrator evaluation system with respect to the following: obtaining useful feedback, supporting improvements to professional practice, informing policy decisions, and adapting to the COVID-19 pandemic?

Most survey respondents reported that their district's evaluation system was "Extremely Actionable" (33.4%) or "Somewhat Actionable" (55.9%). Conversely, 10.6% reported that their evaluative feedback was "Not at All Actionable." One focus group participant described their constructive feedback as follows:

> I feel personally like there's a lot of support. So it's not just an evaluator coming in and criticizing what you are doing. They are giving constructive feedback . . . They give suggestions and tools and are constantly having a conversation of how we can change if it's needed . . . I just feel like they're very involved and helpful on their follow-through and their feedback.

Compared to educators in cities (24.0%), a higher proportion of educators in other areas received "Extremely Actionable" feedback (35.3% for suburb, 41.0% for town, and 35.7% for rural). With respect to tenure status, a larger proportion of nontenured teachers (38.3%) than tenured teachers (28.3%) described their feedback as "Extremely Actionable."

RQ4. What challenges have stakeholders experienced while implementing components of the teacher and administrator evaluation system with respect to the following: obtaining useful feedback, supporting improvements to professional practice, informing policy decisions, reported biases or inequities in system components, and adapting to the COVID-19 pandemic)?

Educators expressed confidence in their understanding of their district's evaluation systems, but also expressed concerns about the evaluation process being burdensome and time consuming.

Most survey respondents reported feeling confident in their understanding of their district's evaluation system, with 52.1% feeling "Extremely Confident," and 39.2% feeling "Somewhat Confident," and the remaining 8.7% feeling "Not Confident" in their understanding of their district's evaluation system. However, during focus groups, several administrators and teachers expressed concerns that teacher evaluations can be overly burdensome, even given some benefits to the process. For example, one focus group participant shared, "This model is so stressful and it's so time consuming and it doesn't feel useful, and I thought I'd heard that originally it was created for teachers to be self-reflecting not for principals to judge."

Administrators' large evaluation caseloads, coupled with an influx of new teachers, the extensiveness of district evaluation processes, and the amount of time the evaluation process takes to complete, were cited by several evaluators and those being evaluated as reasons for viewing the process as a matter of compliance, rather than an opportunity to improve teaching and learning. For example, one educator commented, "I think that's the biggest challenge is that how much time can both sides really invest in it? . . . [S]ome teachers and sometimes administrators can kind of take the mindset of we just got to get it done, check the box."

Educators expressed that the evaluation system was generally accurate, but there were some systematic differences based on their race/ethnicity, their district's geographic locale, and the socioeconomic background of students in their schools.

Most survey respondents viewed the evaluation system's measurement of their professional practice as generally accurate, with 39.5% describing it as "Extremely Accurately" and 50% as "Somewhat Accurately." However, 10.5% of respondents indicated that the system measured their professional practice "Not Accurately at All," indicating some level of concern about accuracy. During focus groups, several teachers, administrators, and joint committee members pointed to the challenge of consistency of evaluations. One joint committee member shared,

> [T]he implementation of that and what that looks like varies from building to building, which is the challenge based upon administrative experience training. Not all of our administrators are trained or retrained by the same people or in the same way. So that variety, even if a building experiences a couple of different administrators in a certain year, a teacher could have a different model or way of being evaluated.

Educators' perceptions of the system's accuracy differed by race/ethnicity and geographic locale. A higher percentage of Black or African American educators (14.8%) than White educators (10.0%) indicated that the system measured their professional practice "Not Accurately At All." The percentage of educators who rated system measurement "Not Accurately at All" was higher among educators in city schools (18.8%) than among educators in suburban (8.5%), town (8.1%), and rural (6.5%) schools. Similarly, a lower percentage of CPS educators (20.3%) reported that their professional practice was measured "Not Accurately at All" than educators in other districts (8.6%).

Educators' views of the evaluation system's accuracy also varied with respect to the socioeconomic background of the students in their schools. Educators in schools within the 1st quartile of low-income student enrollment (i.e., with fewer low-income students) were more likely (44.2%) to describe being measured by the system as "Extremely Accurately" compared to their counterparts in schools within the 4th quartile (32.2%).

Although educators generally characterized the evaluation's formal observation component as a fair measure of their professional practice, there were considerable differences between groups of educators defined by their race/ethnicity, locale, and some raised concerns about subjectivity, bias, and scheduling challenges.

Most survey respondents described how fairly they were measured by the formal observation component as "Extremely Fairly" (50.8%) or "Somewhat Fairly" (42.0%).[24] Only 7.2% reported that the formal observation component measured their professional practice "Not Fairly At All." Educators' views on the fairness of the observation component differed based on their race/ethnicity. A lower proportion of Black or African American (30.6%) than White (53.1%) educators reported that the observation component reflected their professional practice "Extremely Fairly." Views of fairness also varied with respect to locale. A considerably lower proportion of educators in city schools reported that the observation component measured their professional practice "Extremely Fairly" (35.3%) compared to educators in suburban (54.9%), rural (56.1%), and town (58.4%) schools. Similarly, a lower proportion of CPS educators described how well the observation component measured their practice as "Extremely Fairly" (30.2%) compared to educators in all other districts (55.5%).

Several interview and focus group participants described their concerns about subjectivity, bias, and inconsistency in the application of observation rubrics. As one teacher from a suburban district explained:

> I have seen both sides of the spectrum. I evaluated for 4 years and then started in a new district. So you start all over again with your tenure and everything else . . . Anyway, I guess consistency is key and it's not consistent. What one evaluator thinks is an "Excellent" could be what another evaluator thinks is a "Proficient," even though there's descriptions [on the observation rubric].

Additional issues detailed by educators and evaluators alike were logistical in nature, involving the scheduling of evaluations and timelines associated with the evaluation process. For example, one educator reported that, "We're supposed to be putting things in there and we were trying to find ways to organize everything, but just the physical amount of time that it took me to pull things together so that I could prove that I was doing my job . . . is very frustrating." Evaluators also recounted challenges in managing many observations, with one interviewee sharing, "[T]rying to find kind of the middle ground to make an administrator's job as easy as possible to know as few rubrics as possible, yet still respect the individual jobs that various teachers are doing is really hard . . . that's been probably the biggest challenge." These administrative challenges further complicated the implementation of the evaluation system.

[24] As a reminder, the survey used "formal observation component" as a synonym for "professional practice component." Because survey respondents might not have considered these to be synonyms, we report survey findings using the language respondents encountered.

Educators' perceptions on the fairness of the student assessment component differed substantially based on factors such as the educators' roles and the baseline ELA and math proficiency of their students.

Educators expressed mixed views on the fairness of the student assessment component as a measure of their professional practice, with 34.9% describing it as "Extremely Fairly" compared to 48.8% who characterized it as "Somewhat Fairly."[25] However, 16.3% believed the student assessment component measured their professional practice "Not Fairly at All," suggesting that educators held concerns about potential biases in the tests and their alignment with educational practice. As one special education teacher who participated in a focus group explained:

> I think for us [special education teachers], obviously we would like the students with IEPs [individualized education plans] to grow at [a faster rate] to catch up more or however you want to say it. But the reason why they have IEPs is because they are behind and they are learning at a different rate sometimes depending on the skill especially. And standardized tests don't show if the student has grown in the areas that their IEPs are, but I do still look at those scores for indications of things I need to work on and hope to see them grow. But I think when it comes to using that as a measure of whether or not I'm doing my job, that doesn't go together really well.

Survey respondents across different roles, locales, and schoolwide student proficiency rates reported differing perceptions regarding the fairness of the student assessment component as a measure of their professional practice. In the survey sample, general education teachers were less inclined to describe how well the student assessment component reflected their practice as "Extremely Fairly" (28.0%) compared to their colleagues who taught elective subjects (42.8%). Additionally, a considerably lower proportion of city educators (22.5%) rated their measurement by student assessment component as "Extremely Fairly" compared to rural educators (43.7%). CPS educators (22.9%) were less likely than those in all other Illinois districts (36.9%) to describe the student assessment component's measurement as "Extremely Fairly." Furthermore, educators in schools within the 1st (and lowest) quartile of student ELA (29.8%) and math (30.9%) proficiency were less likely to rate the student assessment component as "Extremely Fairly" compared to their peers in the 4th (and highest) quartile of student ELA (46.1%) and math (42.9%) proficiency.

[25] As a reminder, the survey used "student assessment component" as a synonym for the "student growth component." Because survey respondents might not have considered these to be synonyms, we report survey findings using the language respondents encountered.

Several educators expressed concerns about the burden and manipulability of SLOs.

During focus groups, several administrators and teachers enumerated specific challenges with SLOs. Some educators reported that the SLOs were susceptible to manipulation. For example, one administrator shared:

> SLO has nothing to do with teacher instruction. So many times, over the last dozen years it's been, I'm going to give them a test, they're going to take it. They're not going to do well. I'm going to teach 'em the stuff. I'm going to give them the test, they're going to take it, they're going to do well. I'm going to get my 30% excellent and that's going to I'm done. Check, check, check. It's a task.

Some teacher focus group participants shared the sentiment around SLO manipulation that one teacher from a city district described:

> I've gone from the ground up twice now in SLOs—once at my previous district and then at the district that I'm at now. And both have felt very much like a box you just have to check and unfortunately not terribly meaningful because you can manipulate the data any which way you want to.

Additionally, several teachers and administrators who participated in focus groups expressed that the time required to develop SLOs does not justify the investment, reporting that it neither significantly aids teachers' professional development nor markedly enhances student learning. As one administrator from a suburban district shared, "SLOs are ineffective, and yet we spend time on it and so that's time that we could do other more effective things."

Some teachers highlighted increased stress and strained relationships with evaluators as significant challenges associated with evaluation.

Several teachers who participated in focus groups named challenges related to the affective dimensions of evaluation, including increased job-related stress, worry, anxiety, and, in some cases, diminished trust and collegiality between evaluators and teachers. For example, one interviewee explained,

> I think [the evaluation system] messed with the school climate. I think it messed with some of the positive stuff. I think it messed with administrators not sure of what they're supposed to do, teachers aren't sure of what they're supposed to expect. And so you create that uncertainty and anytime you create that uncertainty, it just makes for a different level of chaos in a school that's unnecessary. I mean, so it messes with the trust factors.

Some focus group participants highlighted challenges related to the COVID-19 pandemic.

Although not a commonly identified challenge, some focus group participants noted there were disruptions in the cadence of the evaluation cycle for tenured teachers as a result of the COVID-19 pandemic. Additionally, several districts opted to assign a uniform "Proficient" rating for student growth during the height of the pandemic (2019–20, 2020–21 school years). This decision, intended to address pandemic-related disruptions to schooling, inadvertently caused some tenured teachers who previously received "Excellent" ratings in student growth to experience a decrease in their overall summative evaluation scores.

RQ5. How do stakeholders suggest addressing reported challenges? What supports, resources, and changes are recommended?

Educators—particularly in city schools or schools with lower rates of student ELA proficiency in 2021–22—reported mixed experiences regarding the difficulty of obtaining support, and recommended enhancing the system by providing informal, nonevaluative observations, offering frequent coaching, and focusing on a few specific domains of the observation rubric to foster improvement.

Although getting support to improve practice was considered "Not Difficult at All" (57.9%) or "Somewhat Difficult" (27.7%) by the majority of respondents, a large group described it as "Extremely Difficult" (14.4%). Views differed among respondents based on the locale of their schools and the baseline ELA proficiency rates of the school's students. For example, a substantially higher percentage of educators in city schools (27.8%) reported getting support was "Extremely Difficult" than educators in suburban (11.6%), town (9.0%), and rural (2.9%) schools. Similarly, 22.4% of educators in schools in the 1st (lowest) quartile of student baseline ELA proficiency reported that getting support was "Extremely Difficult"; 8.9% and 9.2% of educators in schools in the 3rd and 4th (highest) quartiles of student baseline ELA proficiency, respectively, agreed.

To better align the evaluation system with opportunities for professional growth, several focus group participants suggested that districts should provide instructional coaches to informally observe teachers and provide nonevaluative feedback. As one city administrator explained:

> So one of the benefits we have is we do have two instructional coaches within our school, and the instructional coaches are nonevaluative, and they are there just to help the teachers. And so as a [department] chair [and evaluator], I can refer them to an instructional coach, but I cannot mandate that they go to an instructional coach, which is good. It's a very healthy relationship where I say, 'Hey, you've been needing some help with classroom management. I'd recommend you go see the instructional coach,

> and that conversation is private and it doesn't get back to me' . . . So it's a very teacher-centered, positive environment.

Several administrators and teachers in the focus groups recommended instituting more frequent, nonevaluative observations of professional practice. This approach, they explained, aims to create a more humanizing, authentic, and relational improvement process that is not solely focused on evaluation. To address the time and workload barrier to providing support, some teachers and administrators suggested that schools or districts focus their efforts on a few specific domains of the observation rubric.

Educators suggested increased attention to relational elements, such as how teacher-evaluator relationships can support more specific or holistic feedback.

Although most survey respondents found their evaluators to be "Extremely Knowledgeable" (50.8%) or "Somewhat Knowledgeable" (40.8%), a notable contingent felt their evaluators were "Not at All Knowledgeable" (8.7%). During focus group discussions, teachers suggested assigning evaluators with content and/or grade-level expertise to help remediate this concern.

During focus groups, educators described how some evaluators have fostered positive relationships with teachers by providing holistic and specific feedback. For example, as one administrator shared:

> So we've had to kind of take off our evaluator hat at times and [say], 'I'm not speaking to you as an evaluator. I want you to grow. I want to see your practice growing. I want to see you successful. So right now, I'm giving you advice that I would give myself as a teacher.' So sometimes you almost have to humanize that part of teaching to be able to see through that lens too and have that empathy for your teachers. Because I know they're all trying their best.

This intentionality in fostering the teacher-evaluator relationship is also captured by a joint committee member:

> I feel personally like there's a lot of support. So it's not just an evaluator coming in and criticizing what you are doing. They are giving constructive feedback . . . They give suggestions and tools and are constantly having a conversation of how we can change if it's needed . . . So I think it's helpful creating that relationship between evaluator and teacher.

In cases where observations are intentionally less evaluative or punitive, teachers and administrators both expressed that the evaluation process can help strengthen professional practice and relationships.

RQ6. What are the distributions of educator ratings across districts and schools and over time, and to what extent do educator ratings vary systematically with respect to the characteristics of educators, districts, and schools?

Of the educators who received an evaluation rating in 2022–23, more than 97% of teachers and more than 96% of principals and assistant principals received a rating of "Excellent" or "Proficient."

Of the 138,639 teachers in the data, 37.4% received an evaluation rating in 2022–23. Nearly all teachers who received a rating (97.2%) were rated "Excellent" or "Proficient." Only 2.6% of teachers who received a rating were rated "Needs Improvement," and 0.2% were rated "Unsatisfactory." Of the 7,024 principals and assistant principals in the data, 32.0% received an evaluation rating in 2022–23. Of these, only 3.2% received ratings of "Needs Improvement" or "Unsatisfactory." The other 96.8% were rated either "Excellent" or "Proficient."

The distribution of evaluation ratings varied by educator race/ethnicity, as well as by school and district characteristics.

Among educators who received a rating in 2022–23, the percentages of educators who were rated "Excellent" or "Proficient" were lower among Black or African American and Hispanic or Latinx educators, educators with fewer years of experience, and educators in schools and districts with fewer resources.

Specifically, only 22.1% of Black or African American teachers and 32.0% of Hispanic or Latinx teachers who received a rating were rated "Excellent" compared to 50.0% of White teachers; by contrast, 10.6% of Black or African American teachers and 5.5% of Hispanic or Latinx teachers who received a rating were rated "Needs Improvement" or "Unsatisfactory" compared to only 2.0% of White teachers. Black or African American and Hispanic or Latinx teachers consistently received lower ratings than White teachers, regardless of whether they had 1 year, 2–3 years, 4–5 years, or 6 or more years of teaching experience. For instance, among first-year teachers, 10.8% of Black or African American teachers and 15.0% of Hispanic or Latinx teachers were rated as "Excellent," compared with 24.0% of White teachers. Conversely, 12.6% of Black or African American teachers and 9.3% of Hispanic or Latinx teachers were rated as "Needs Improvement" or "Unsatisfactory," compared with 4.6% of White teachers.

Higher percentages of teachers at schools with a large proportion of low-income student enrollment received "Excellent" ratings, and lower percentages received "Needs Improvement"

or “Unsatisfactory” ratings than their peers in schools with smaller low-income student enrollment. Specifically, teachers who received a rating at schools in the 4th (highest) quartile of low-income student enrollment had lower percentages of receiving an “Excellent” rating (26.3%) than their peers who received a rating at schools in the 1st (lowest) quartile of low-income student enrollment (55.4%). Similarly, a higher percentage of teachers who received a rating at schools in the 4th quartile of low-income student enrollment received “Needs Improvement” or “Unsatisfactory” (8.3%) ratings than their peers in the 1st quartile (1.2%) who taught fewer low-income students.

The distribution of evaluation ratings also differed by the district’s teacher characteristics. During a focus group discussion, one member of a joint committee conveyed that Black or African American teachers working in high-poverty schools may encounter complex barriers to implementing an evaluation system that both adequately and consistently reflects the true effectiveness of educators working in such contexts:

> I would add, in terms of challenges, there've been several studies that have shown that African American teachers in general score lower on [their evaluations] than White teachers. And that's totally related to the fact that more Black teachers choose to work in high-poverty schools where the complexities of the problems, the education problems, are of course intertwined with the problems of poverty and [our evaluation system] just doesn't capture whether somebody's a good teacher or not in those situations, or it doesn't do it all the time.

An early contributor to PERA echoed this concern during an interview:

> I hear my colleagues in Chicago talk about the Black and Brown issues that teachers up there are facing. And we know from the data that Black and Brown teachers have many more negative evaluations . . . It's difficult for me to believe that those educators are that much worse than their fellow educators across the building or across the state. I think there are definitely still issues of bias, and I'm not sure what the source is, but I do know that there are issues of fairness across the state.

Evaluation ratings varied depending on the characteristics of teachers in each district. Teachers who received ratings in districts with larger proportions of White teachers tended to receive higher ratings (38.0% of Quartile 1 were rated “Excellent” versus 57.7% for Quartile 4). A higher percentage of teachers who received ratings were rated “Excellent” in districts with more experienced teachers (61.8% for Quartile 4 compared to 37.5% for Quartile 1). Similarly, a higher percentage of teachers who received ratings were rated “Excellent” in districts with higher teacher retention rates (63.2% for Quartile 4 versus 35.0% for Quartile 1).

Additionally, compared to their White counterparts, Black or African American and Hispanic or Latinx principals and assistant principals who received ratings had substantially lower rates of receiving "Excellent" ratings (27.0% and 20.0% versus 51.2%) and higher rates of receiving "Needs Improvement" or "Unsatisfactory" (7.9% and 10.4% versus 2.0%) ratings. Principals and assistant principals with 6 or more years of experience who received ratings had much higher rates of receiving "Excellent" ratings (56.5%) compared to first-year principals and assistant principals (27.9%). Only 22.4% of principals and assistant principals who received ratings and worked in schools in the 4th quartile of low-income student enrollment were rated "Excellent"; by contrast, 54.0% of principals and assistant principals in schools in the 2nd quartile received "Excellent" ratings. Principals and assistant principals who received ratings in districts with higher average teacher experience (52.7% for Quartile 3 versus 31.1% for Quartile 1) and higher district teacher retention rates (50.1% for Quartile 3 versus 30.9% for Quartile 1) tended to have higher rates of receiving an "Excellent" rating.

Potential Recommendations for Consideration

PEAC is charged with developing recommendations to continuously improve the policy and implementation of educator evaluation using the findings from this study (Illinois General Assembly, 2024-b). Based on these integrated findings, AIR identified the following potential recommendations for PEAC to consider to inform a set of policy modifications and statewide supports that address any challenges identified while preserving the core benefits of PERA. These potential policy recommendations have been grouped into two categories: potential recommendations related to the implementation of Article 24A of the Illinois School Code (Illinois General Assembly, 2024-a) and potential recommendations related to revisions to Article 24A of the Illinois School code itself.

Potential Recommendations Related to the Implementation of Article 24A of the Illinois School Code:

- Minimize the burden of the evaluation system while prioritizing actionable feedback and professional growth for all educators.
- Calibrate evaluators and multiple observers.
- Support diverse and culturally inclusive evaluation practices.
- Foster positive relationships and trust between administrators and teachers.

Potential Policy Recommendations Related to Revisions to Article 24A of the Illinois School Code:

- Reconsider the use or incorporation of student growth.

Potential Recommendations Related to the Implementation of Article 24A of the Illinois School Code:

Minimize the Burden of the Evaluation System While Prioritizing Actionable Feedback and Professional Growth for All Educators.

Educators in the focus groups shared concerns that educator evaluation is often time consuming, and the quality or amount of feedback or support may not be worth the time investment for every aspect of the evaluation system. In focus groups, educators shared that some components of the evaluation system, such as SLOs, were overly burdensome, and the process did not provide educators with actionable feedback that would be worth the time invested. However, some respondents shared that they would be interested in more frequent informal observations or walk-throughs that could provide more opportunities for more frequent feedback. PEAC may consider recommendations to prioritize components of the evaluation system that offer greater opportunities for actionable feedback, such as classroom observations, over more time-intensive processes, such as SLOs, which may not provide educators with actionable feedback.

In developing recommendations, PEAC may want to consider supports that will help Illinois to prioritize actionable feedback and professional growth, so that evaluators and educators are more likely to view the system as an opportunity to enhance teaching and learning, rather than viewing the process as a matter of compliance. For example, to address the discrepancy of coaching and feedback opportunities between tenured and nontenured educators, PEAC may consider recommending developing or modifying guidance and training for school leaders on how to strengthen pre- and postconferences. These conferences could be leveraged as another coaching opportunity for all educators. School leaders could use the postconference meeting to reflect on the observation with the teacher, collectively brainstorm on what could be improved, and discuss what supports the teacher would need to apply the instructional practices discussed. Strengthening pre- and postconferences could help address the discrepancy in coaching received by tenured teachers and help to connect the evaluation system more clearly to improving teachers' instructional practice. This builds a coaching opportunity into an existing meeting without additional time-burden.

Another way to strengthen the connection between the evaluation system and feedback is to streamline and target evaluation components to strategic district priorities or the professional needs of educators. Examples of how to prioritize specific components of professional practice frameworks like the Danielson Framework for Teaching based on district priorities or educator needs could be developed and made available to all districts. For example, the Massachusetts Department of Elementary and Secondary Education developed focus indicators to help

prioritize high-leverage practices and streamline the formal evaluation process (Massachusetts Department of Elementary and Secondary Education, 2023).

Calibrate Evaluators and Support Multiple Observers.

Currently, evaluators must undergo training and retraining to become evaluators. The analysis of evaluation ratings data identified systematic differences in educator ratings by the race and ethnicity of teachers, which suggests that evaluators may need additional calibration. One potential consideration is to revisit the cadence and content of evaluator training to reduce subjectivity and bias and improve interrater reliability among evaluators. For example, to become a certified CLASS observer requires a 2-day training and passing an online assessment that involves watching and coding five 20-minute videos of classroom practice. Observers are required to complete an annual recertification that includes coding additional videos. Existing evaluator training could be strengthened by including video examples highlighting potential areas of bias or subjectivity.

To help provide more frequent feedback opportunities, PEAC may consider recommending guidance and support for districts to offer multiple observers, including teacher leaders. This could provide multiple perspectives on a teacher's practice and give some flexibility to allow educators with the same content or subject area as the teacher being observed to observe and give feedback. Training additional observers might also help ease the time burden for administrators, and it could provide teachers with feedback specific to their content or subject area.

Support Diverse and Culturally Inclusive Evaluation Practices.

In the evaluation ratings data, educator evaluation ratings varied across racial and ethnic lines, and in the survey data, concerns regarding the fairness and accuracy of the evaluation system differed substantially by race/ethnicity and locale. This may suggest that the implementation of the evaluation system could be continuously improved by embedding more diverse and culturally inclusive evaluation practices in the system by, for example, developing “look-fors” that illustrate examples of instructional practices from a variety of cultures. The set of culturally responsive look-fors created by Massachusetts for their model teacher evaluation rubric (Massachusetts Department of Elementary and Secondary Education, n.d.) presents one such example. Embedding principals of diversity, equity, and inclusion in training for evaluators or other professional learning opportunities could be another approach to supporting diverse and culturally inclusive evaluation practices.

PEAC may also consider recommending that ISBE help address concerns of bias in the evaluation system by regularly reporting on teacher evaluation ratings by teacher and school characteristics in a way that also protects teacher privacy. Producing annual tables similar to

those presented in the Educator Evaluation Ratings Patterns and Trends section of this report may help alleviate perceptions of bias or spur action to address any identified bias from districts, the general assembly, or others in the education community.

Foster Positive Relationships and Trust Between Administrators and Teachers.

Establishing trust is essential for any coaching relationship between teachers and school leaders (Bryk & Schneider, 2002; Cranston, 2011; Tschannen-Moran, 2001). In focus groups, some teachers expressed increased strain on the relationship between teachers and evaluators. Several teachers and administrators highlighted relationships as a key part of successful evaluation systems, which entailed informal coaching and walk-throughs. One consideration for PEAC is a statewide and/or regional professional learning community focused on strategies to build trust and positive relationships between teachers and evaluators with opportunities to share examples for different building sizes. Broadly, a professional learning community in this form could include areas such as the following:

- Using evaluation data to inform coaching and educator professional development overall.
- Communicating the connection between evaluation system data and professional learning needs in the district and/or school.
- Implementing strategies to improve school culture and working conditions.
- Incorporating educator voice in decision making and the implementation of the evaluation system and professional learning.

Potential Recommendations Related to Revisions to Article 24A of the Illinois School Code:

Reconsider the Use or Incorporation of Student Growth.

This report's review of district evaluation plans and focus groups with teachers and school leaders suggest that most districts in Illinois use SLOs to measure student growth. Reconsidering the state's approach to student growth is a resonant theme culled from data. There is little, if any, evidence that SLOs are a valid, reliable measure of teacher effectiveness. Educators shared that student growth, particularly SLOs, were perceived as unfair, easily manipulated, and time consuming to implement. Some examples of how the state might change its use of student growth measures to evaluate teachers include the following:

- Eliminate the student growth requirement from evaluations.
- Reduce the percentage weight of student growth in the overall evaluation score.

- Offer a statewide student-growth model, such as a VAM or SGP, for teachers of Grades 4–8 ELA and math, and eliminate the student growth requirement for all other teachers.[26] There is evidence that a well-designed VAM in Grades 4–8 ELA and math can provide a valid measure of teacher contributions to student learning, and a statewide model could reduce the time burden on teachers of measures like SLOs. However, a high-quality statewide VAM could not provide a growth rating to teachers in all grades and subjects, and it may be unfair to have evaluation requirements for some teachers and not others.
- Embed the instructional practices from the SLO process—such as analyzing student data, setting growth targets, and using data to inform instruction—into the professional practice rubric. This would eliminate the separate weighted measure of student growth but maintain the emphasis on the data analytic practices that inform instruction. This approach could also include a focus on the impact on student learning within the professional practice framework. An example of this approach is Ohio's Teacher Performance Evaluation Rubric (Ohio Department of Education, 2020), where SLO processes are evident in components such as use of high-quality student data, communication with students, use of assessments, and evidence of student learning.

The state of Illinois commissioned this evaluation of PERA implementation to inform potential policy recommendations and improve the overall evaluation system for educators across the state. This study provided a broad set of evidence, including a policy scan of teacher evaluation systems across the United States, a survey of educators who have experienced PERA implementation, interviews and focus groups with early contributors to legislation as well as educators involved in its implementation, a review of a wide cross section of district evaluation plans, and an analysis of educator ratings across the state. The potential policy recommendations presented here are based on the findings from a synthesis of these data and are intended to inform PEAC's recommendations regarding further changes to the overall teacher evaluation system in the state of Illinois.

[26] Most of the research on the validity and reliability of using student growth to measure teacher contributions to student learning has focused on VAMs of student achievement in ELA and math in Grades 4–8 (e.g., Bacher-Hicks et al., 2014; Chetty et al., 2014; Glazerman et al., 2013). This research suggests that the bias of VAMs in Grades 4–8 ELA and math is small (Koedel et al., 2015). In Grades 4–8 ELA and math, students typically have a test score from the prior grade that has a strong curricular and correlational relationship with the current year's test score, and the VAMs use that prior-year test score to control, in part, for nonrandom sorting of students into schools and classrooms.

References

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, 333(6045), 1034–1037. https://doi.org/10.1126/science.1207998

American Educational Research Association. (2015). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher, 44*(8), 448–452.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

American Statistical Association. (2014). *ASA statement on using value-added models for educational assessment*. https://www.amstat.org/asa/files/pdfs/POL-ASAVAM-Statement.pdf

Bacher-Hicks, A., Kane, T. J., & Staiger, D. O. (2014). *Validating teacher effect estimates using changes in teacher assignments in Los Angeles.* (NBER Working Paper No. 20657). National Bureau of Economic Research. https://www.nber.org/papers/w20657

Bacher-Hicks, A., Chin, M. J., Kane, T. J., & Staiger, D. O. (2019). An experimental evaluation of three teacher quality measures: Value-added, classroom observations, and student surveys. *Economics of Education Review*, *73*(C), 101919.

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., Rothstein, R., Shavelson, R. J., & Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers* (Briefing Paper #278). Economic Policy Institute.

Barone, C. (2017). What ESSA says: Continuities and departures. In F.M. Hess & M. Eden (Eds.), *The Every Student Succeeds Act: What it means for schools, systems, and states* (pp. 59–74). Havard Education Press.

Bartanen, B. (2020). Principal quality and student attendance. *Educational Researcher*, *49*(2), 101–113.

Bill and Melinda Gates Foundation. (2012). *Asking students about teaching: Student perception surveys and their implementation.* https://usprogram.gatesfoundation.org/news-and-insights/usp-resource-center/resources/asking-students-about-teaching-student-perception-surveys-and-their-implementation

Bowers, A. J., & White, B. R. (2014). Do principal preparation and teacher qualifications influence different types of school growth trajectories in Illinois? A growth mixture model analysis. *Journal of Educational Administration, 52*(5), 705-736.

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

Castellano, K. E., & Ho, A. D. (2013). *A practitioner's guide to growth models.* Council of Chief State School Officers.

Charmaz, K. (2008). Grounded theory as an emergent method. In S. N. Hesse-Biber & P. Leavy (Eds.), *Handbook of emergent methods* (pp. 155–172). Guilford. http://www.sxf.uevora.pt/wp-content/uploads/2013/03/Charmaz_2008-b.pdf

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review, 104*(9), 2593-2632.

Close, K., Amrein-Beardsley, A., & Collins, C. (2020). Putting teacher evaluation systems on the map: An overview of states' teacher evaluation systems post-Every Student Succeeds Act. *Education Policy Analysis Archives*, *28*(58).

Colorado Department of Education. (2023). *State model evaluation system for teachers*. https://www.cde.state.co.us/educatoreffectiveness/smes-teacher#teacherrubric

Connally, K., & Tooley, M. (2016). *Beyond ratings: Re-envisioning state teacher evaluation systems as tools for professional growth.* New America.

Council of Chief State School Officers. (2013). Interstate teacher assessment and support consortium In *TASC model core teaching standards and learning progressions for teachers 1.0: A resource for ongoing teacher development*. https://ccsso.org/sites/default/files/2017-12/2013_INTASC_Learning_Progressions_for_Teachers.pdf

Cranston, J. (2011). Relational trust: The glue that binds a professional learning community. *Alberta Journal of Educational Research, 57*(1), 59–72. https://doi.org/10.11575/ajer.v57i1.55455

Czarniawska, B. (2008). Organizing: How to study it and how to write about it. *Qualitative Research in Organizations and Management: An International Journal*, *3*(1), 4–20. https://doi.org/10.1108/17465640810870364

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Association for Supervision and Curriculum Development.

Denzin, N., & Lincoln, Y. L. (2003). *Collecting and interpreting qualitative materials* (2nd ed.). Sage.

Dey, I. (1993). *Qualitative data analysis: A user-friendly guide for social scientists*. Routledge.

Doss, C. J. (2019). *Student growth percentiles 101: Using relative ranks in student test scores to help measure teaching effectiveness.* RAND Corporation. https://www.rand.org/pubs/research_reports/RR4312z2.html.

Ferguson, R. F., & Danielson, C. (2015). How framework for teaching and tripod 7cs evidence distinguish key components of effective teaching. In *Designing teacher evaluation systems* (pp. 98–143). https://doi.org/https://doi.org/10.1002/9781119210856.ch4

Garet, M. S., Wayne, A. J., Brown, S., Rickles, J., Song, M., & Manzeske, D. (2017*). The impact of providing performance feedback to teachers and principals*. NCEE 2018-4001. National Center for Education Evaluation and Regional Assistance. https://files.eric.ed.gov/fulltext/ED578873.pdf

Geiger, T. J., & Amrein-Beardsley, A. (2017). Administrators gaming test- and observation-based teacher evaluation methods: To conform to or confront the system [Article]. *AASA Journal of Scholarship & Practice*, *14*, 45+.

Geiger, T., & Amrein-Beardsley, A. (2019). Student perception surveys for K–12 teacher evaluation in the United States: A survey of surveys. *Cogent Education, 6*(1), 1602943. https://doi.org/10.1080/2331186X.2019.1602943

Geverdt, D., (2018). *Education demographic and geographic estimates (EDGE) Geocodes: Public schools and local education agencies, 2016-2017* (NCES 2018-080). U.S. Department of Education. National Center for Education Statistics. http://nces.ed.gov/pubsearch/

Glazerman, S., Protik, A., Teh, B., Bruch, J., Max, J. (2013). *Transfer incentives for high-performing teachers: Final results from a multisite experiment* (NCEE 2014-4003). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Goe, L., Biggers, K., & Croft, A. (2012). *Linking teacher evaluation to professional development: Focusing on improving teaching and learning*. National Comprehensive Center for Teacher Quality.

Goldhaber, D., & Hansen, M. (2013). Is it just a bad class? Assessing the long-term stability of estimated teacher performance. *Economica*, *80*(319), 589–612.

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher*, *44*(2), 96–104.

Greene, J. C. (2008). Is mixed methods social inquiry a distinctive methodology? *Journal of Mixed Methods Research*, *2*(1), 7–22. https://doi.org/10.1177/1558689807309969

Guarino, C., Reckase, M., Stacy, B., & Wooldridge, J. (2015). A comparison of student growth percentiles and value-added models of teacher performance. *Statistics and Public Policy, 2*(1), 1–11.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, *18*(1), 59–82. https://doi.org/10.1177/1525822X05279

Hannah, D. R., & Lautsch, B. A. (2011). Counting in qualitative research: Why to conduct it, when to avoid it, and when to closet it. *Journal of Management Inquiry*, *20*(1), 14–22. https://doi.org/10.1177/1056492610375

Hunter, S. B. (2022). High-leverage teacher evaluation practices for instructional improvement. *Educational Management Administration & Leadership.* https://doi.org/10.1177/17411432221112

Illinois Association of Regional School Superintendents. (2015). *Map: Regional offices of education*. https://iarss.org/map/

Illinois General Assembly. (2010). Public Act 96-0861 [Senate Bill 315] Performance Evaluation Reform Act. https://www.ilga.gov/legislation/publicacts/fulltext.asp?Name=096-0861

Illinois General Assembly. (2024-a). Illinois Compiled Statutes. Schools, 105 ILCS 5/24A. Evaluation of Certified Employees. https://www.ilga.gov/legislation/ilcs/ilcs4.asp?DocName=010500050HArt.+24A&ActID=1005&ChapterID=17&SeqStart=168900000&SeqEnd=170300000

Illinois General Assembly. (2024-b). Illinois Compiled Statutes. Schools, 105 ILCS 5/21B-20. Evaluation of Certified Employees. https://www.ilga.gov/legislation/BillStatus.asp?DocNum=463&GAID=17&DocTypeID=SB&SessionID=112&GA=103

Illinois General Assembly. (2022-a). Title 23 Illinois Administrative Code Part 50: Evaluation of Educator Licensed Employees Under Articles 24A and 34 of the School Code. https://www.ilga.gov/commission/jcar/admincode/023/02300050sections.html

Illinois General Assembly. (2022-b). Title 23 Illinois Administrative Code Part 24: Standards for All Illinois Educators. https://www.ilga.gov/commission/jcar/admincode/023/02300024sections.html

Illinois General Assembly. (2022-c). Senate Resolution 0774. https://www.ilga.gov/legislation/BillStatus.asp?DocNum=774&GAID=16&DocTypeID=SR&LegId=138641&SessionID=110&GA=102

Illinois State Board of Education. (2013). *Guidance document 13–06: Model teacher evaluation system measuring student growth using type III assessments.* https://www.isbe.net/Documents_PEAC/13-6-te-model-sys-meas-typeiii.pdf

Illinois State Board of Education. (2015). *Non-regulatory guidance. The Performance Evaluation Reform Act (PERA) and Senate Bill (SB7).* https://www.isbe.net/Documents/pera_guidance.pdf

Illinois State Board of Education. (2023). *Illinois report card data.* Retrieved July 24, 2023, from https://www.isbe.net/ilreportcarddata

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, *18*(1), 3–20. https://doi.org/10.1177/1525822X05282

Jackson, C. K. (2012). *Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina* (Working Paper 18624). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w18624/w18624.pdf

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy, 126*(5), 2072–2107. https://doi.org/10.1086/699018

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, *26*(1), 101–136.

Jacob, A., & McGovern, K. (2015). *The mirage: Confronting the hard truth about our quest for teacher development.* TNTP. https://files.eric.ed.gov/fulltext/ED558206.pdf

Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation.* (NBER Working Paper No. 14607). National Bureau of Economic Research. http://www.nber.org/papers/w14607

Kane, T. J. (2024). Develop and validate—Then scale. *Education Next*, 24(2). https://www.educationnext.org/develop-validate-scale-lessons-gates-foundation-effective-teaching-strategy/

Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review, 47*(2015), 180–195.

Kraft, M. A., & Gilmour, A. F. (2017). Revisiting the Widget Effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher 46*(5), 234–239.

Kuhfeld, M. (2017). When students grade their teachers: A validity analysis of the Tripod student survey. *Educational Assessment, 22*(4), 253–274.

LeCompte, M. D. (2000). Analyzing qualitative data. *Theory Into Practice, 39*(3), 146–154. https://doi.org/10.1207/s15430421tip3903_5

Lin, S., Luo, W., Tong, F., Irby, B. J., Alecio, R. L., Rodriguez, L., & Chapa, S. (2020). Data-based student learning objectives for teacher evaluation. *Cogent Education*, *7*(1), 1713427.

Lovison, V., & Taylor, E. S. (2018). Can teacher evaluation programs improve teaching? (*Getting Down to Facts II report).* Stanford University Center for Education Policy Analysis. https://gettingdowntofacts.com/sites/default/files/2018-09/GDTFII_Report_Taylor.pdf

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319–383). Springer.

Martínez, J. F., Schweig, J., & Goldschmidt, P. (2016). Approaches for combining multiple measures of teacher performance: Reliability, validity, and implications for evaluation policy. *Educational Evaluation and Policy Analysis*, *38*(4), 738-756.

Marzano, R. (2017). *The Marzano focused teacher evaluation model*. https://www.marzanocenter.com/wp-content/uploads/sites/4/2018/10/MC06-14-FTEM-White-Paper-1-16-18-Digital-4.pdf

Marzano, R. J., Carbaugh, B., Rutherford, A., & Toth, M. D. (2013a). *Marzano Center teacher observation protocol for the 2014 Marzano teacher evaluation model.* Marzano Center/Learning Sciences International. https://sde.ok.gov/sites/default/files/documents/files/2014%20Protocol%20Paper_20140128.pdf

Marzano, R. J., & Toth, M. D. (2013b). *Teacher evaluation that makes a difference: A new model for teacher growth and student achievement.* Association for Supervision and Curriculum Development.

Massachusetts Department of Elementary and Secondary Education. (n.d.) *Culturally responsive look-fors*. https://www.doe.mass.edu/edeval/resources/calibration/look-fors.docx

Massachusetts Department of Elementary and Secondary Education. (2023). *Focus indicators for teaching in 2023–2024.* https://www.doe.mass.edu/edeval/implementation/teachers-focus-indicators.pdf

Maxwell, J. A. (2013). *Qualitative research design: An interactive approach*. Sage.

Maxwell, J. A. (2021). Why qualitative methods are necessary for generalization. *Qualitative Psychology, 8*(1), 111–118. https://doi.org/10.1037/qup0000173

Michigan Department of Education. (2023). *What is a student growth percentile*? https://www.michigan.gov/mde/-/media/Project/Websites/mde/OEAA/Accountability/General/Student_Growth_Percentiles.pdf?rev=65965ed157294013a957eaa2ba25506d&hash=5C1A2561AF8429A9CC2DF5C77C101306

Milanowski, A., Ristow, L., Finster, M., McKittrick, C., Doli, M., Lewandowski, H., Roseland, D., White, B. R., Carl, B., Gawade, N., Kang, H., Marlin, D., Meyer, R., Ponisciak, S., Steele, C., Wang, Y. (2016). *Evaluation of the Illinois Performance Evaluation Reform Act: Final report*. Westat & Wisconsin Center for Education Research. https://www.isbe.net/documents/pera-final-report-160630.pdf

Missouri Department of Elementary and Secondary Education (2023). Model evaluation system. https://dese.mo.gov/educator-quality/educator-growth-toolbox/model-evaluation-system

Mizrav, E. (2019). Decision points: Defining, calculating, and addressing gaps in access to effective educators. Considerations for states. Special issues brief. *Center on Great Teachers and Leaders*.

Morganstein, D., & Wasserstein, R. (2014). ASA statement on value-added models. *Statistics and Public Policy, 1*(1), 108–110.

National Center for Education Statistics. (2023a). *Common Core of Data: Public elementary/secondary school universe survey, 2021–22, provisional version 1a*. https://nces.ed.gov/ccd/files.asp

National Center for Education Statistics. (2023b). *Common Core of Data: Local education agency universe survey, 2021–22, provisional version 1a*. https://nces.ed.gov/ccd/files.asp

National Council on Teacher Quality. (2021–22). State Teacher Policy Database; Wis. Stat. § 115.415(2a). https://www.nctq.org/yearbook/home

National Council on Teacher Quality. (2022). *Teacher and principal evaluation policies*. https://www.nctq.org/publications/State-of-the-States-2022:-Teacher-and-Principal-Evaluation-Policies

Ohio Department of Education. (2020). *Teacher performance evaluation rubric*. https://dam.assets.ohio.gov/image/upload/sboe.ohio.gov/Educator-Evaluations/TeacherEvaluations/Rubric-with-Editable-Evidence-Row.pdf

Opper, I.M. (2019). *Value-added modeling 101: Using student test scores to help measure teaching effectiveness.* RAND Corporation. https://www.rand.org/pubs/research_reports/RR4312z1.html

Papay, J. P., Taylor, E. S., Tyler, J. H., & Laski, M. E. (2020). Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data. *American Economic Journal: Economic Policy*, *12*(1), 359-388.

Patton, M. Q. (2002). *Qualitative research & evaluation methods*. Sage.

Pivovarova, M., Amrein-Beardsley, A., & Geiger T. (2016). *Value-added models: What the experts say.* Kappan. https://kappanonline.org/value-added-models-what-the-experts-say/

Reform Support Network. (2014, May). *A toolkit for implementing high-quality student learning objectives 2.0.* https://www.education-first.com/wp-content/uploads/2015/10/RSN-SLO-Toolkit.pdf

Ross, E., & Walsh, K. (2019). *State of the states 2019: Teacher and principal evaluation policy.* National Council on Teacher Quality.

Seastrom, M. M. (2010). Statistical methods for protecting personally identifiable information in aggregate reporting (NCES No. 2011-603). *National Center for Education Statistics SLDS Technical Brief*, *3*. https://nces.ed.gov/pubs2011/2011603.pdf

Shaha, S. H., Glassett, K. F., & Copas, A. (2015). The impact of teacher observations with coordinated professional development on student performance: A 27-state program evaluation. *Journal of College Teaching & Learning, 12*(1), 55.

StataCorp. (2023). *Stata statistical software: Release 18*.

Steinberg, M. P., & Garrett, R. (2016). Classroom composition and measured teacher performance: What do teacher observation scores really measure? *Educational Evaluation and Policy Analysis*, *38*(2), 293-317.

Steinberg, M. P., & Sartain, L. (2021). What explains the race gap in teacher performance ratings? Evidence from Chicago Public Schools. *Educational Evaluation and Policy Analysis*, *43*(1), 60-82.

Swisher, A., & Saenz-Armstrong, P. (2022). *State of the states 2022: Teacher and principal evaluation policies*. National Council on Teacher Quality.

Tashakkori, A., & Creswell, J. W. (2007). Editorial: The new era of mixed methods. *Journal of Mixed Methods Research*, *1*(1), 3–7. https://doi.org/10.1177/2345678906293042

Thomas, D. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation, 27*(2), 237–246. https://doi.org/10.1177/1098214005283748

Tschannen-Moran, M. (2001). Collaboration and the need for trust. *Journal of Educational Administration, 39*(4), 308–331.

Tuma, A. P., Hamilton, L. S., & Tsai, T. (2018). *How do teachers perceive feedback and evaluation systems? Findings from the American teacher panel.* RAND.

U.S. Department of Education. (2023). *Illinois State Board of Education State Template for the Consolidated State Plan Under the Every Student Succeeds Act*. https://www.isbe.net/Documents/Approved-Illinois-State-Plan-2023.pdf

Walsh, E., & Isenberg, E. (2015). How does value added compare to student growth percentiles? *Statistics and Public Policy*, *2*(1), 1–13.

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. TNTP.

White, M., & Maher, B. L. (2024). How might rubric-based observations better support teacher learning and development? *Educational Research*, *66*(1), 86–101.

Whitehurst, G., Chingos, M. M., & Lindquist, K. M. (2014). *Evaluating teachers with classroom observations.* Brown Center on Education Policy: Brookings Institute.

Yeh, S. S. (2013). A reanalysis of the effects of teacher replacement using value-added modeling. *Teachers College Record*, *115*(12), 1–35.

Yin, R. K. (2006). Mixed methods research: Are the methods genuinely integrated or merely parallel? *Research in the Schools*, *13*(1), 41–47.

Appendix A. Survey Questionnaire

Exhibit A–1. Survey Questionnaire

Illinois State Board of Education Performance Evaluation Reform Act Implementation Survey

1. Were you evaluated in an Illinois public school as a tenured or nontenured teacher, principal, assistant principal, or other school-based administrator during the previous school year (2022–23)? Please select one: (required)
   a. Yes
   b. No (Go to thank you page.)
2. How **frequently** was your professional practice **observed** as part of your district's evaluation system during the previous school year (2022–23)? Please select one:
   a. Never
   b. Once
   c. Twice
   d. Three times
   e. Four or more times
   f. Unsure
3. How **frequently** did you engage in a **preconference** before you were observed as part of your district's evaluation system during the previous school year (2022–23)? Please select one:
   a. Never
   b. Once
   c. Twice
   d. Three times
   e. Four or more times
   f. Unsure
4. How **frequently** did you engage in a **postconference** after you were observed as part of your district's evaluation system during the previous school year (2022–23)? Please select one:
   a. Never
   b. Once
   c. Twice
   d. Three times
   e. Four or more times
   f. Unsure
5. How **frequently** did you receive **written feedback** based on an observation as part of your district's evaluation system during the previous school year (2022–23)? Please select one:
   a. Never (Go to Q7.)
   b. Once
   c. Twice

d. Three times
e. Four or more times
f. Unsure (Go to Q7.)

(If no response, go to Q7.)

6. How **actionable was the written and verbal postconference feedback** that you received as part of your district's evaluation system during the previous school year (2022–23)? Please select one:
   a. Not at all actionable
   b. Somewhat actionable
   c. Extremely actionable
   d. Unsure

7. How **frequently** were you **offered coaching** as a result of an observation as part of your district's evaluation system during the previous school year (2022–23)? Please select one:
   a. Never (Go to Q9.)
   b. Once
   c. Twice
   d. Three times
   e. Four or more times
   f. Unsure (Go to Q9.)

(If no response, go to Q9.)

8. How **useful** was the **coaching** you received as part of your district's evaluation system during the previous school year (2022–23)? Please select one:
   a. Not at all useful
   b. Somewhat useful
   c. Extremely useful
   d. Unsure

9. Overall, how **supportive** was the district's evaluation system to your professional growth and development during the previous school year (2022–23)?
   a. Not at all supportive
   b. Somewhat supportive
   c. Extremely supportive
   d. Unsure

10. How confident did you feel in your **understanding** of the district's evaluation system during the previous school year (2022–23)? Please select one:
    a. Not confident
    b. Somewhat confident
    c. Extremely confident
    d. Unsure

11. Which **components** did the district's evaluation system use to rate your professional practice during the previous school year (2022–23)? Please select all that apply:
    a. Standardized, statewide assessments (e.g., Illinois Assessment of Readiness, Illinois Science Assessment, SAT)
    b. Districtwide or locally developed student assessments, including student learning objectives
    c. Observation rubrics for professional practice (e.g., Framework for Teaching or Classroom Assessment Scoring System) (If c is selected and neither a nor b is selected, go to Q13.)
    d. Other measures of educator practice (Please specify:) (If only d and/or e is selected, go to Q14.)
    e. Unsure (If only d and/or e is selected, go to Q14.)

(If no response, go to Q14.)

12. How **fairly** did the **student assessment component** of the district's evaluation system measure your professional practice during the previous school year (2022–23)? Please select one:
    a. Not at all fairly
    b. Somewhat fairly
    c. Extremely fairly
    d. Unsure

13. (Q11 = "observation rubrics for professional practice") How **fairly** did the **formal observation component** of the district's evaluation system measure your professional practice during the previous school year (2022–23)? Please select one:
    a. Not fairly at all
    b. Somewhat fairly
    c. Extremely fairly
    d. Unsure

14. How **accurately** did the district's evaluation system measure the effectiveness of your overall professional practice during the previous school year (2022–23)? Please select one:
    a. Not at all accurately
    b. Somewhat accurately
    c. Extremely accurately
    d. Unsure

15. How **knowledgeable** was your **evaluator** about your professional practice during the previous school year (2022–23)? Please select one:
    a. Not at all knowledgeable
    b. Somewhat knowledgeable
    c. Extremely knowledgeable
    d. Unsure

16. Think of the hours of professional learning you accessed or received during the previous school year (2022–23). How many of these hours, if any, were directly connected to feedback you received as part of your evaluation? Please select one:
    a. None
    b. Less than 1 hour
    c. 1–2 hours
    d. 3–5 hours
    e. 6–10 hours
    f. 11 or more hours
    g. Unsure

17. If you requested **extra support** (e.g., informal observations, informal feedback, or informal mentorship from colleagues) to improve your practice, how **difficult** was it to get the support that you needed during the previous school year (2022–23)? Please select one:
    a. Not at all difficult
    b. Somewhat difficult
    c. Extremely difficult
    d. Unsure
    e. I did not request additional support.

18. What was your **role** during the previous school year (2022–23)? If you worked in more than one role, please select the role in which you spent most of your time. Please select one:
    a. Nontenured teacher
    b. Tenured teacher
    c. Principal
    d. Assistant principal
    e. Other school-based administrator
    f. Related service provider (e.g., school counselor or nurse)

(If no response, go to Q24.)

19. At the conclusion of the prior school year (2022–23), how many **years of experience** did you have in this role? Please select one:
    a. 2022–23 was my first year in this role.
    b. 2022–23 was my second or third year in this role.
    c. 2022–23 was my fourth or fifth year in this role.
    d. 2022–23 was my sixth or more year in this role.

20. (If Q18 = "tenured teacher," "principal," "assistant principal," "other school-based administrator," or "related service provider") Did you serve as an **evaluator** during the previous school year (2022–23)? Please select one:
    a. Yes
    b. No

c. Unsure

(Go to Q24 if Q18 = "principal," "assistant principal," "other school-based administrator," or "related service provider.")

21. What **grade levels** did you work with during the previous school year (2022–23)? Please select all that apply:
    a. Early childhood (PK–2)
    b. Elementary (3–5)
    c. Middle (6–8)
    d. High (9–12)

22. What **subject area(s)** did you teach during the previous school year (2022–23)? Please select all that apply:
    a. General education
    b. Special education
    c. English language arts
    d. Mathematics
    e. Science
    f. History/social science
    g. Visual/performing arts
    h. Physical/health education
    i. Foreign language
    j. Career and technical education
    k. Other (Please specify:)

23. During the previous school year (2022–23), did you teach a **grade and subject tested** by the Illinois statewide student assessment system (e.g., Illinois Assessment of Readiness, Illinois Science Assessment, SAT)? Please select one:
    a. Yes
    b. No
    c. Unsure

24. Were teachers in your district represented by any of the following organizations during the previous school year (2022–23)? Please select all that apply:
    a. Illinois Education Association
    b. Illinois Federation of Teachers
    c. Chicago Teachers Union
    d. Other
    e. Unaffiliated
    f. I don't know/prefer not to answer.

25. Which of the following options best describes your **race or ethnicity**?[27] Please select all that apply:
    a. Hispanic/Latinx
    b. White
    c. Black/African American
    d. Asian
    e. American Indian/Alaska Native
    f. Native Hawaiian/Pacific Islander
    g. Other/I don't know (Please specify:)

26. Are you: (Please select all that apply:)
    a. Male?
    b. Female?
    c. Transgender, nonbinary, or another gender?

27. Are you interested in sharing more about your experiences with the district's evaluation system during a 60-minute virtual **focus group** with the American Institutes for Research (AIR) research team in the coming months? Please select one:
    a. Yes
    b. No

28. "Thank you" page.

[27] Race/ethnicity categories are from the U.S. Census Bureau (n.d.), updated annually.

Appendix B. Survey Methodologies

Sample Selection

In this section, the process for selecting a sample of teachers and administrators to participate in the survey is described. In survey research, the *population* is the entire group about which the study seeks to draw conclusions, whereas the *sample* is a subset of that population surveyed to make inferences about the population. Sample selection is a crucial step in designing a survey because it helps ensure that the sample of survey respondents is representative of the full population. The population for this study is all teachers and administrators in districts and schools that implemented the Performance Evaluation Reform Act (PERA) evaluation systems during the 2022–23 school year.

To identify the districts and schools that implemented the PERA evaluation systems in 2022–23, the American Institutes for Research® (AIR®) used administrative data from the 2022–23 Illinois Report Card (Illinois State Board of Education, 2023) supplemented by data from the 2021–22 Common Core of Data (CCD; National Center for Education Statistics, 2023a, 2023b). Based on these data, it was determined that the population would include approximately 136,378 teachers and administrators located across 3,825 schools in 854 districts.[28] Rather than survey the entire population, a representative sample of this population was selected. A survey of a representative sample allows for the analysis of collected data and the generalization of conclusions to the full population while minimizing the cost of the study and the burden placed on the full population of teachers and administrators.

District Selection

If districts were sampled entirely at random, it is possible that, merely by chance, the sampling would result in a set of districts that are not representative of the full population of districts in the state. To avoid this, stratified systematic sampling was used (Cochran, 1977) to select 125 districts that are representative of the full population for the study.

Specifically, five distinct steps to select the 125 districts for the study were followed:

1. Districts were categorized according to their locale, designating them as either “urban,” “suburban,” or “town/rural” based on each district’s locale code as reported in the CCD. We identified 29 urban districts, 343 suburban districts, and 482 town/rural districts in Illinois. Given the limited number of urban districts, it was necessary to include the full population of

[28] Only regular local school districts, as defined by NCES, are included in this count of 854 districts. The NCES definitions of district types can be found here: https://nces.ed.gov/ccd/elsi/glossary.aspx?app=tableGenerator&term=21104.

urban districts in the sample to make confident inferences about the differences in evaluation systems between urban and nonurban districts.[29]

2. The remaining suburban and town/rural districts were grouped into six regional office of education (ROE) service areas, as defined by the Illinois Association of Regional Superintendents of Schools (2023).
3. Within the six ROE service areas, districts were categorized into two groups based on the percentage of students eligible for free or reduced-price lunch (FRPL). A district was categorized as "low FRPL" if the percentage of students from low-income backgrounds was less than or equal to the statewide median percentage of such students for Illinois districts (less than or equal to 40.6%). A district was categorized as "high FRPL" if the percentage of students from low-income backgrounds exceeded the median for districts in Illinois (more than 40.6%). This process resulted in 12 groups of districts (i.e., defined by the six ROE service areas and two FRPL categorizations).
4. Within these 12 groups, districts were further sorted based on the following characteristics:
    a. Districts were sorted based on locale: suburban or town/rural.
    b. Within each urbanicity group, districts were sorted based on student enrollment: small (less than 1,000 students), medium (more than 1,000 but less than or equal to 2,500 students), or large (more than 2,500 students).
    c. Within each urbanicity and enrollment group, districts were sorted based on the percentage of students in the district who are White: low (less than the statewide median percentage of White students for districts in Illinois) or high (greater than the statewide median).
    d. Within each group defined by urbanicity, enrollment, and the percentage of students in the district who are White, districts were sorted based on the percentage of students in the district who are Black or African American: low (less than the statewide median percentage of Black or African American students for districts in Illinois) or high (greater than the statewide median).
    e. Within each group defined by urbanicity, enrollment, and the percentages of students in the district who are White and Black or African American, districts were sorted based on the percentage of students in the district who are Hispanic: low (less than the statewide

[29] Because urban districts typically have more teachers and administrators than nonurban districts, a sample that includes all 37 urban districts will include a disproportionate share of teachers and administrators from urban districts. To address this disproportionality, the analysis will "weight" districts based on how well they represent the broader population of districts in the state. Survey analyses that apply analysis weights to reflect selection probabilities can produce unbiased estimates for the full population.

median percentage of Hispanic students for districts in Illinois) or high (greater than the statewide median).

f. After following these steps, 12 groups of districts (defined by the six ROE service area and two FRPL categorizations) were identified. Within each group, districts were sorted based on five characteristics (urbanicity, enrollment, and the percentages of students who are White, Black or African American, and Hispanic).

5. To complete the final step in selecting 125 districts for the study, districts were randomly sampled from within all 12 groups. Specifically, within each group of districts, a starting point was identified at random and then every *n*th district was selected, where *n* was the ratio of the group's population size to the group's sample size.[30]

This stratified systematic sampling approach helped ensure that the sample is representative of the population in terms of district characteristics that may correlate with the characteristics of the districts' evaluation systems or teachers' and administrators' perceptions of their districts' evaluation systems.

0 compares the characteristics of the 125 districts in the sample with the 854 districts with active enrollment in the broader population based on the 2021–22 Common Core of Data and ISBE 2022 Report Card Public Data available at the time of sampling. For example, 48.0% of the districts in the sample are in ROE Service Area 1, whereas 35.8% of the districts in the population are in ROE Service Area 1. For each characteristic, the selection rate is the number of districts in the sample expressed as a percentage of districts in the population. For example, our sample includes 19.6% of ROE Service Area 1 districts in the population.[31]

[30] The sample size for each stratum is proportional to the square root of the number of teachers and administrators in all districts in the stratum. Counts of the number of teachers and administrators in each district were drawn from the 2022–23 Illinois Report Card (ISBE, 2023).

[31] The characteristics of districts included in the sample do not always reflect the broader population. For example, 56.7% of the districts in the population are characterized as small (i.e., with fewer than 1,000 students) compared with 46.4% of districts (i.e., 58 of 125) in the sample. To address this disproportionality, the analysis will "weight" districts based on how well they represent the broader population of districts in the state. Survey analyses that apply analysis weights to reflect selection probabilities can produce unbiased estimates of the responses of the full population.

Exhibit B–1. Characteristics of Districts in the Full Population and in the Sample

District Characteristics	Number of Districts in the Population (*n* = 854)	Percentage of Districts in the Population (*n* = 854)	Number of Districts in the Sample (*n* = 125)	Percentage of Districts in the Sample (*n* = 125)	Selection Rate
ROE Service Area					
1	306	35.8%	60	48.0%	19.6%
2	120	14.1%	14	11.2%	11.7%
3	109	12.8%	13	10.4%	11.9%
4	112	13.1%	16	12.8%	14.3%
5	84	9.8%	9	7.2%	10.7%
6	123	14.4%	13	10.4%	10.6%
Urbanicity (district level)					
Urban	29	3.4%	29	23.2%	100.0%
Suburban	343	40.2%	49	39.2%	14.3%
Town/rural	482	56.4%	47	37.6%	9.8%
FRPL (district level)					
Low (≤ 40.6%)	427	50.0%	62	49.6%	14.5%
High (> 40.6%)	427	50.0%	63	50.4%	14.8%

Note. Total percentages may be affected by rounding error. ROE = regional office of education. FRPL is the percentage of students in a district who are eligible to participate in the federal free or reduced-price lunch program. The median district-level percentage of students eligible for FRPL is 40.6%.

School Selection

A stratified systematic sampling approach was used to select 300 schools from the 125 districts as follows:

- It was determined that the number of schools selected from each district should be proportional to the square root of the number of teachers and administrators in the district, while guaranteeing that at least one school per district is in the sample. This ensured that each district was assigned a fair share of the 300 schools, and at least one school was selected from each district.
- The process used to sort schools is analogous to the process used to sort districts. Specifically, schools were sorted by their Title I status, urbanicity, the percentage of students from low-income backgrounds, the percentage of White students, the percentage of Black or African American students, and the percentage of Hispanic students.

- Finally, a random starting point was identified and every *n*th school within each district was selected until the predetermined number of schools was identified.

0 compares the characteristics of the 300 schools we selected versus all 3,825 schools in the population. For example, 20.7% of the schools in the sample are urban schools compared with 25.4% of the schools in the population.[32] For each characteristic, the selection rate is the number of schools in the sample expressed as a percentage of schools in the population. For example, our sample includes 6.4% of the urban schools in the population.

Exhibit B–2. Characteristics of Schools in the Full Population and in the Sample

School Characteristics	Number of Schools in the Population (*n* = 3,841)	Percentage of Schools in the Population (*n* = 3,841)	Number of Schools in the Sample (*n* = 300)	Percentage of Schools in the Sample (*n* = 300)	Selection Rate
ROE service area					
1	2,131	55.7%	152	50.7%	7.1%
2	394	10.3%	37	12.3%	9.4%
3	375	9.8%	26	8.7%	6.9%
4	367	9.6%	39	13.0%	10.6%
5	294	7.7%	23	7.7%	7.8%
6	264	6.9%	23	7.7%	8.7%
Urbanicity (school level)					
Urban	970	25.4%	62	20.7%	6.4%
Suburban	1,583	41.4%	124	41.3%	7.8%
Town/rural	1,272	33.3%	114	38.0%	9.0%
FRPL (school level)					
Low (≤ 46.2%)	1,913	50.0%	177	59.0%	9.3%
High (> 46.2%)	1,912	50.0%	123	41.0%	6.4%

Note. Total percentages may be affected by rounding error ROE = regional office of education. FRPL is the percentage of students in a district who are eligible to participate in the federal free or reduced-price lunch program. The median school-level percentage of students eligible for FRPL is 46.2%.

[32] In cases in which the sample and population percentages of schools substantially differ, the analysis will "weight" schools based on how well they represent the broader population of schools in the selected districts. Survey analyses that apply analysis weights to reflect selection probabilities can produce unbiased estimates of the responses of the full population.

Teacher and Administrator Selection

All 11,393 teachers and administrators within selected schools were invited to take the survey. 0 presents the characteristics of teachers and administrators in the 300 selected schools and the full population. For example, 45.6% of the teachers and administrators in the population work in high-FRPL schools, but 36.3% of the teachers and administrators in the sample work in high-FRPL schools.[33]

Exhibit B–3. Characteristics of the Teachers' Schools in the Full Population and in the Sample

School Characteristics	Number of Teachers in the Population (*n* = 136,378)	Percentage of Teachers in the Population (*n* = 136,378)	Number of Teachers in the Sample (*n* = 11,393)	Percentage of Teachers in the Sample (*n* = 11,393)
ROE service area				
1	91,696	67.2%	7,656	67.2%
2	11,013	8.1%	883	7.8%
3	9,546	7.0%	634	5.6%
4	9,897	7.3%	1,076	9.4%
5	8,382	6.1%	576	5.1%
6	5,844	4.3%	567	5.0%
Urbanicity (school level)				
Urban	37,278	27.3%	2,790	24.5%
Suburban	68,640	50.3%	5,884	51.6%
Town/rural	30,460	22.3%	2,719	23.9%
FRPL (school level)				
Low (≤ 46.2%)	74,253	54.4%	7,252	63.7%
High (> 46.2%)	62,125	45.6%	4,141	36.3%

Note. Total percentages may be affected by rounding error. ROE = regional office of education. FRPL is the percentage of students in a district who are eligible to participate in the federal free or reduced-price lunch program. The median school-level percentage of students eligible for FRPL is 46.2%.

[33] In cases in which the sample and population percentages of schools substantially differ, our analysis will "weight" teachers based on how well they represent the broader population of teachers in the state. Survey analyses that apply analysis weights to reflect selection probabilities can produce unbiased estimates of the responses of the full population.

Appendix C. Survey Administration

The survey was administered to teachers and school-based administrators in a representative sample of Illinois schools whose districts implemented the Performance Evaluation Reform Act (PERA) in 2022–23. A survey of a representative sample, rather than a census survey, was planned for the study because a sample survey creates a greater likelihood of identifying accurate conclusions about the population while minimizing the burden placed on Illinois educators and the logistical challenges of data collection.

After the sample of schools was identified, the survey team was able to obtain access to the list of teachers and administrators with demographic information and email addresses in all public schools in noncharter districts in the state of Illinois. Survey invitations were sent to all 11,709 teachers and administrators in the 298 sampled schools across 124 districts. Prenotifications were sent to sample members via email on November 3, 2023, followed by the first survey invitations on November 7, 2023. Several follow-up emails were sent throughout the survey window to nonrespondents between November 13, 2023, and December 15, 2023. The survey was closed on December 15, 2023. The survey contact activities are shown in 0.

Exhibit C–1. Survey Contact Activities

Activity	Date
Email prenotification	November 3, 2023
Initial email invitation	November 7, 2023
Email follow-up 1	November 13, 2023
Email follow-up 2	November 20, 2023
Email follow-up 3	November 28, 2023
Email follow-up 4	December 1, 2023
Email follow-up 5	December 5, 2023
Email follow-up 6	December 11, 2023
Email follow-up 7	December 13, 2023
Final email reminder	December 15, 2023

Appendix D. Survey Eligibility and Response Rates

As shown in Exhibit D–1, responses were submitted from 3,967 sampled educators in 294 schools across 124 districts, but 1,172 of them were not eligible for the survey because they were not evaluated in an Illinois public school during the 2022–23 school year. Thus, there is an eligibility rate of 70.5% (1,172/3,967). Note that educators might not be formally rated in official evaluations, but still choose to participate in the survey because they might consider themselves as being evaluated if they received any type of evaluations during the 2022–23 school year. In addition, because only 33.9% (3,967/11,709) of sampled educators answered the survey, there were potential risks of nonresponse bias in the analytic sample even if the eligibility issue was addressed.

An eligible respondent of the survey as a sampled educator was defined as being one who provided a valid response (i.e., not missing or "unsure") to any of the questions about their experience with their districts educator evaluation systems and supports, and indicated they were evaluated in an Illinois public school during the 2022–23 school year. Based on this definition, there were 2,588 usable responses from educators in 121 districts and 289 schools. Using American Association of Public Opinion Research Response Rate 3 (AAPOR RR3) (AAPOR, 2023), the response rate for the survey was

$$AAPOR\ RR3 = \frac{C}{C + P + e * U} = \frac{2{,}588}{2{,}588 + 207 + 0.7046 * (11{,}709 - 3{,}967)} = 31.4\%$$

where C is complete, P is partial complete, e is eligibility rate, and U is unknown eligibility.

Exhibit D–1. Disposition of Sampled Educators

Disposition of Sampled Educators	Number of Educators
Who responded to the survey	3,967
Who answered "yes" to the eligibility question	2,795
Complete	2,588
Partially complete	207
Who did not respond to the survey	7,742
Total	11,709

Survey Bias Analysis

As described above, there were ineligible educators in the set of sampled educators, and fewer than a third of sampled educators provided usable responses to the survey. If the analysis were

done using the set of respondents without any adjustments to the sampling weights, there might be bias in the results because there were differences between sampled educators and those who answered the survey, and between the set of sampled educators and those who were eligible and provided usable information to the survey, as shown in Exhibit D–2.

Between sampled educators and those who answered the survey, there were statistically significant differences in the percentages of those who were teachers or administrators and those who had been in the school for 6 or more years. Between sampled educators and those who were eligible and provided usable information to the survey, there were statistically significant differences in the percentages of those who were teachers or administrators, those who were White, and those who were in the school for 1 year or less, 2 or 3 years, and 6 or more years.

To adjust for these differences resulting from the differential eligibility and nonresponse patterns, the sampling weights were adjusted to reduce the potential bias in weighted analysis, as described in the next section.

Exhibit D–2. Percentage of Teachers in the Full Sample, in the Sample of Educators Who Responded to the Survey, and in the Sample of Educators Who Provided Usable Responses to the Survey, Weighted by Sampling Weights

Variable	Category	% Full Sample (*n* = 11,709)	% Who Responded to the Survey (*n* = 3,967)	% Who Provided Usable Responses to the Survey (*n* = 2,588)
Region	1	65.5%	59.9%	62.9%
Region	2	10.6%	11.3%	10.9%
Region	3	5.6%	6.1%	5.6%
Region	4	7.6%	8.3%	8.4%
Region	5	6.6%	9.6%	8.4%
Region	6	4.2%	4.8%	3.9%
School Urbanicity	City	22.1%	18.0%	20.7%
School Urbanicity	Suburban	53.4%	52.0%	51.5%
School Urbanicity	Town	10.2%	12.9%	12.3%
School Urbanicity	Rural	14.2%	17.1%	15.6%
School Level	Elementary	55.6%	52.6%	53.9%
School Level	Middle	19.1%	19.6%	18.6%

Variable	Category	% Full Sample (*n* = 11,709)	% Who Responded to the Survey (*n* = 3,967)	% Who Provided Usable Responses to the Survey (*n* = 2,588)
School Level	High	25.2%	27.8%	27.4%
School in Chicago Public Schools	No	85.2%	87.7%	84.9%
School in Chicago Public Schools	Yes	14.8%	12.3%	15.1%
School Enrollment Quartile: 2022–23	First (Percentiles 1–25)	12.6%	13.8%	13.5%
School Enrollment Quartile: 2022–24	Second (Percentiles 26–50)	17.9%	18.4%	18.6%
School Enrollment Quartile: 2022–25	Third (Percentiles 51–75)	26.6%	26.8%	26.2%
School Enrollment Quartile: 2022–26	Fourth (Percentiles 76–99)	42.9%	41.0%	41.6%
School Low-Income Enrollment Quartile: 2022–23	First (Percentiles 1–25)	32.0%	31.2%	29.1%
School Low-Income Enrollment Quartile: 2022–24	Second (Percentiles 26–50)	17.7%	21.1%	20.0%
School Low-Income Enrollment Quartile: 2022–25	Third (Percentiles 51–75)	26.5%	25.1%	23.8%
School Low-Income Enrollment Quartile: 2022–26	Fourth (Percentiles 76–99)	23.9%	22.7%	27.0%
School ELA Proficiency Rate Quartile: 2022–23	First (Percentiles 1–25)	28.6%	28.2%	33.1%
School ELA Proficiency Rate Quartile: 2022–24	Second (Percentiles 26–50)	23.1%	23.4%	19.9%
School ELA Proficiency Rate Quartile: 2022–25	Third (Percentiles 51–75)	27.0%	27.6%	27.0%
School ELA Proficiency Rate Quartile: 2022–26	Fourth (Percentiles 76–99)	21.3%	20.8%	20.0%

Variable	Category	% Full Sample (*n* = 11,709)	% Who Responded to the Survey (*n* = 3,967)	% Who Provided Usable Responses to the Survey (*n* = 2,588)
School Mathematics Proficiency Rate Quartile: 2022–23	First (Percentiles 1–25)	30.2%	30.3%	33.9%
School Mathematics Proficiency Rate Quartile: 2022–24	Second (Percentiles 26–50)	20.1%	18.8%	17.7%
School Mathematics Proficiency Rate Quartile: 2022–25	Third (Percentiles 51–75)	22.9%	24.2%	23.9%
School Mathematics Proficiency Rate Quartile: 2022–26	Fourth (Percentiles 76–99)	26.9%	26.8%	24.5%
Role	Principal	2.9%	**4.0%**	**5.4%**
Role	Teacher	97.1%	**96.0%**	**94.6%**
Race/Ethnicity	Black or African American	9.7%	7.8%	8.5%
Race/Ethnicity	American Indian or Alaska Native	0.1%	–	–
Race/Ethnicity	Asian	1.7%	1.8%	2.3%
Race/Ethnicity	Hispanic or Latino	5.9%	5.2%	6.4%
Race/Ethnicity	Native Hawaiian or Other Pacific Islander	0.0%	–	–
Race/Ethnicity	White	80.1%	83.0%	80.2%
Race/Ethnicity	Two or More Races & Unknown	2.4%	1.9%	2.3%
Race/Ethnicity	Male	22.7%	21.1%	21.1%
Race/Ethnicity	Female	77.3%	78.9%	78.9%
Years in School in Current Position	1 year	16.2%	15.4%	**19.4%**
Years in School in Current Position	2–3 years	17.0%	15.8%	**20.4%**
Years in School in Current Position	4–5 years	13.7%	12.2%	12.4%

Variable	Category	% Full Sample (*n* = 11,709)	% Who Responded to the Survey (*n* = 3,967)	% Who Provided Usable Responses to the Survey (*n* = 2,588)
Years in School in Current Position	6 or more years	53.1%	**56.5%**	**47.7%**
Grades PK–2	No	49.1%	53.5%	51.9%
Grades PK–2	Yes	50.9%	46.5%	48.1%
Grades 3–5	No	47.8%	51.6%	50.4%
Grades 3–5	Yes	52.2%	48.4%	49.6%
Grades 6–8	No	35.2%	38.4%	38.2%
Grades 6–8	Yes	64.8%	61.6%	61.8%
Grades 9–12	No	74.8%	72.2%	72.6%
Grades 9–12	Yes	25.2%	27.8%	27.4%

Note. The statistics for the full sample were treated as a fixed value in the statistical testing. The statistics that were statistically significant from the corresponding value from the full sample are in **bold**. Statistics are not shown (–) because the group being analyzed is too small. ELA = English language arts.

Appendix E. Survey Weighting Adjustments

To adjust for the differences resulting from the differential eligibility and nonresponse patterns, a multinomial logistic model with educator and school characteristics as predictors to predict the sampled educators' response probability/propensity was estimated. Logistic regression modeling has been used to predict the probability of responding to adjust for nonresponse (Lepkowski et al., 1989). A multinomial logistic model was used because sampled educators could be nonresponsive, answered the survey but did not provide useful information on a district evaluation, or provided useful information on a district evaluation to the survey. A multinomial logistic model allows for the simultaneous estimation of the probability of being in any of the three categories.

Using the estimated probabilities from the model, the sampling weights were adjusted to account for differential response patterns as follows:

$$AW_r = SW/p_r$$

$$AW_{re} = SW/p_{re}$$

where AW_r is the adjusted weight for nonresponse, SW is the sampling weight that is the inverse of selection probability, p_r is the estimated probability of responding to the survey, AW_{re} is the adjusted weight for eligibility and whether respondents provided usable information on a district evaluation, and p_r is the estimated probability of being eligible and providing usable information on a district evaluation. The adjusted weights were then used to recalculate the percentages examined in 0. The new percentages are shown in 0, which shows no significant differences after the weighting adjustments and indicates that the adjustments effectively reduced potential biases in the distributions of the variables examined.

Exhibit E–1. Percentage of Teachers in the Full Sample Weighted by Sampling Weights, in the Sample of Educators Who Responded to the Survey Weighted by Adjusted Weights, and in the Sample of Educators Who Provided Usable Responses to the Survey Weighted by Adjusted Weights

Variable	Category	% Full Sample (*n* = 11,709)	% Who Responded to the Survey (*n* = 3,967)	% Who Provided Usable Responses to the Survey (*n* = 2,588)
Region	1	65.5%	62.8%	64.7%
Region	2	10.6%	11.1%	11.0%
Region	3	5.6%	5.9%	5.5%

Variable	Category	% Full Sample (*n* = 11,709)	% Who Responded to the Survey (*n* = 3,967)	% Who Provided Usable Responses to the Survey (*n* = 2,588)
Region	4	7.6%	7.2%	7.5%
Region	5	6.6%	8.9%	7.9%
Region	6	4.2%	4.0%	3.4%
School Urbanicity	City	22.1%	20.1%	21.0%
School Urbanicity	Suburban	53.4%	55.4%	54.7%
School Urbanicity	Town	10.2%	10.7%	10.7%
School Urbanicity	Rural	14.2%	13.8%	13.5%
School Level	Elementary	55.6%	55.7%	55.8%
School Level	Middle	19.1%	20.0%	20.5%
School Level	High	25.2%	24.3%	23.7%
School in Chicago Public Schools	No	85.2%	86.3%	84.7%
School in Chicago Public Schools	Yes	14.8%	13.7%	15.3%
School Enrollment Quartile: 2022–23	First (Percentiles 1–25)	12.6%	12.6%	12.5%
School Enrollment Quartile: 2022–24	Second (Percentiles 26–50)	17.9%	18.3%	18.9%
School Enrollment Quartile: 2022–25	Third (Percentiles 51–75)	26.6%	27.6%	27.0%
School Enrollment Quartile: 2022–26	Fourth (Percentiles 76–99)	42.9%	41.4%	41.6%
School Low- Income Enrollment Quartile: 2022–23	First (Percentiles 1–25)	32.0%	30.5%	28.4%
School Low- Income Enrollment Quartile: 2022–24	Second (Percentiles 26–50)	17.7%	19.1%	18.7%
School Low- Income Enrollment: Quartile 2022–25	Third (Percentiles 51–75)	26.5%	25.8%	24.4%
School Low- Income Enrollment Quartile: 2022–26	Fourth (Percentiles 76–99)	23.9%	24.5%	28.5%
School ELA Proficiency Rate Quartile: 2022–23	First (Percentiles 1–25)	28.6%	29.2%	33.8%

Variable	Category	% Full Sample (*n* = 11,709)	% Who Responded to the Survey (*n* = 3,967)	% Who Provided Usable Responses to the Survey (*n* = 2,588)
School ELA Proficiency Rate Quartile: 2022–24	Second (Percentiles 26–50)	23.1%	22.2%	19.1%
School ELA Proficiency Rate Quartile: 2022–25	Third (Percentiles 51–75)	27.0%	26.8%	26.9%
School ELA Proficiency Rate Quartile: 2022–26	Fourth (Percentiles 76–99)	21.3%	21.8%	20.2%
School Mathematics Proficiency Rate Quartile: 2022–23	First (Percentiles 1–25)	30.2%	31.2%	34.4%
School Mathematics Proficiency Rate Quartile: 2022–24	Second (Percentiles 26–50)	20.1%	18.3%	17.9%
School Mathematics Proficiency Rate Quartile: 2022–25	Third (Percentiles 51–75)	22.9%	23.3%	23.2%
School Mathematics Proficiency Rate Quartile: 2022–26	Fourth (Percentiles 76–99)	26.9%	27.3%	24.4%
Role	Principal	2.9%	3.2%	3.3%
Role	Teacher	97.1%	96.8%	96.7%
Race/Ethnicity	Black or African American	9.7%	8.2%	8.3%
Race/Ethnicity	American Indian or Alaska Native	0.1%	–	–
Race/Ethnicity	Asian	1.7%	2.0%	2.2%
Race/Ethnicity	Hispanic or Latino	5.9%	5.7%	6.6%
Race/Ethnicity	Native Hawaiian or Other Pacific Islander	0.0%	–	–
Race/Ethnicity	White	80.1%	81.9%	80.4%
Race/Ethnicity	Two or More Races & Unknown	2.4%	1.9%	2.1%
Race/Ethnicity	Male	22.7%	22.2%	21.3%
Race/Ethnicity	Female	77.3%	77.8%	78.7%

Variable	Category	% Full Sample (*n* = 11,709)	% Who Responded to the Survey (*n* = 3,967)	% Who Provided Usable Responses to the Survey (*n* = 2,588)
Years in School in Current Position	1 year	16.2%	17.4%	17.4%
Years in School in Current Position	2–3 years	17.0%	16.5%	16.6%
Years in School in Current Position	4–5 years	13.7%	12.9%	13.1%
Years in School in Current Position	6 or more years	53.1%	53.2%	52.9%
Grades PK–2	No	49.1%	50.7%	50.1%
Grades PK–2	Yes	50.9%	49.3%	49.9%
Grades 3–5	No	47.8%	48.6%	48.7%
Grades 3–5	Yes	52.2%	51.4%	51.3%
Grades 6–8	No	35.2%	35.2%	35.0%
Grades 6–8	Yes	64.8%	64.8%	65.0%
Grades 9–12	No	74.8%	75.7%	76.3%
Grades 9–12	Yes	25.2%	24.3%	23.7%

Note. The statistics for the full sample were treated as a fixed value in the statistical testing. Statistics are not shown (–) because the group being analyzed is too small. ELA = English language arts.

Because 2,795 sampled members answered "yes" to the eligibility question while only 2,588 provided valid responses (and thus are included in the study analytic sample), the adjusted weights for the analytic sample who provided usable responses needed to be further adjusted to account for eligibility, as the adjusted weights for the last column of 0 weight the analytic sample to the original full population. To account for eligibility, the same multinomial logistic model was reestimated, except that in this model the three categories of the dependent variable change to the following: nonresponsive, responded to the survey but answered "no" to the eligibility question, or answered "yes" to the eligibility question. After the adjustments, the weights brought the analytic sample down to the population of eligible educators. At the same time, the population weights (unity) for every educator on the full list of educators whose answers were received were adjusted in order to compare the population distribution of eligible educators and the estimated distributions from the analytic sample. The weights were adjusted as follows:

$$FAW = AW_r * p_{e|r}$$

$$AW_{pop} = 1 * p_{e|r}$$

where FAW is the final adjusted weight for the analytic sample, AW_r is the adjusted weight for nonresponse, $p_{e|r}$ is the estimated probability of being eligible for the survey given that the sampled educator responded to the survey, and AW_{pop} is the adjusted weight for eligibility for the population. This set of final adjusted weights was used in the analysis and accounted for stratification, clustering, and finite population correction in the estimation of variance through linearization using the statistical software Stata's built-in *svyset* command.

As shown in 0, there is only one statistically significant difference (in the percentage of educators who were in the school 1 year or less), which is not a concern given the large number of comparisons made in data analysis. In contrast, there would be seven statistically significant differences if the analytic sample were weighted by the sampling weights directly without any postsurvey adjustments.

Exhibit E–2. Percentage of Eligible Educators in the Population, in the Sample of Eligible Educators Who Provided Valid Responses to the Survey Weighted by Adjusted Weights, and in the Sample of Eligible Educators Who Provided Valid Responses to the Survey Weighted by Sampling Weights

Variable	Category	% Eligible Population (*n* = 98,557)	% Who Provided Valid Responses to the Survey Weighted by Adjusted Weights (*n* = 2,588)	% Who Provided Valid Responses to the Survey Weighted by Sampling Weights (*n* = 2,588)
Region	1	68.1%	65.8%	62.9%
Region	2	7.9%	10.5%	10.9%
Region	3	7.0%	5.4%	5.6%
Region	4	7.2%	7.3%	8.4%
Region	5	5.9%	7.7%	8.4%
Region	6	4.0%	3.2%	3.9%
School Urbanicity	City	29.4%	22.4%	**20.7%**
School Urbanicity	Suburban	49.7%	54.9%	51.5%
School Urbanicity	Town	9.8%	10.1%	12.3%
School Urbanicity	Rural	11.0%	12.5%	15.6%
School Level	Elementary	53.8%	56.6%	53.9%
School Level	Middle	15.1%	19.1%	18.6%
School Level	High	31.1%	24.3%	27.4%

Variable	Category	% Eligible Population (*n* = 98,557)	% Who Provided Valid Responses to the Survey Weighted by Adjusted Weights (*n* = 2,588)	% Who Provided Valid Responses to the Survey Weighted by Sampling Weights (*n* = 2,588)
School in Chicago Public Schools	No	82.0%	83.6%	84.9%
School in Chicago Public Schools	Yes	18.0%	16.4%	15.1%
School Enrollment Quartile: 2022–23	First (Percentiles 1–25)	9.8%	12.2%	13.5%
School Enrollment Quartile: 2022–24	Second (Percentiles 26–50)	17.2%	18.5%	18.6%
School Enrollment Quartile: 2022–25	Third (Percentiles 51–75)	23.9%	27.2%	26.2%
School Enrollment Quartile: 2022–26	Fourth (Percentiles 76–99)	49.1%	42.1%	41.6%
School Low- Income Enrollment Quartile: 2022–23	First (Percentiles 1–25)	30.4%-	28.3%	29.1%
School Low- Income Enrollment Quartile: 2022–24	Second (Percentiles 26–50)	22.4%	18.4%	20.0%
School Low- Income Enrollment Quartile: 2022–25	Third (Percentiles 51–75)	24.3%	24.3%	23.8%
School Low- Income Enrollment Quartile: 2022–26	Fourth (Percentiles 76–99)	22.9%	29.0%	27.0%
School ELA Proficiency Rate Quartile: 2022–23	First (Percentiles 1–25)	25.6%	34.0%	33.1%
School ELA Proficiency Rate Quartile: 2022–24	Second (Percentiles 26–50)	22.8%	18.6%	19.9%
School ELA Proficiency Rate Quartile: 2022–25	Third (Percentiles 51–75)	23.4%	26.7%	27.0%
School ELA Proficiency Rate Quartile: 2022–26	Fourth (Percentiles 76–99)	28.1%	20.8%	**20.0%**

Variable	Category	% Eligible Population (*n* = 98,557)	% Who Provided Valid Responses to the Survey Weighted by Adjusted Weights (*n* = 2,588)	% Who Provided Valid Responses to the Survey Weighted by Sampling Weights (*n* = 2,588)
School Mathematics Proficiency Rate Quartile: 2022–23	First (Percentiles 1–25)	24.6%	34.6%	33.9%
School Mathematics Proficiency Rate Quartile: 2022–24	Second (Percentiles 26–50)	22.2%	17.1%	17.7%
School Mathematics Proficiency Rate Quartile: 2022–25	Third (Percentiles 51–75)	22.9%	23.3%	23.9%
School Mathematics Proficiency Rate Quartile: 2022–26	Fourth (Percentiles 76–99)	30.3%	25.0%	24.5%
Role	Principal	3.5%	4.0%	**5.4%**
Role	Teacher	96.5%	96.0%	**94.6%**
Race/Ethnicity	Black or African American	7.0%	8.7%	8.5%
Race/Ethnicity	American Indian or Alaska Native	0.2%	–	–
Race/Ethnicity	Asian	2.0%	2.4%	2.3%
Race/Ethnicity	Hispanic or Latino	8.9%	7.0%	**6.4%**
Race/Ethnicity	Native Hawaiian or Other Pacific Islander	0.1%	–	–
Race/Ethnicity	White	79.3%	79.2%	80.2%
Race/Ethnicity	Two or More Races & Unknown	2.6%	2.3%	2.3%
Race/Ethnicity	Male	24.8%	22.0%	**21.1%**
Race/Ethnicity	Female	75.2%	78.0%	**78.9%**
Years in School in Current Position	1 year	17.7%	**21.3%**	19.4%
Years in School in Current Position	2–3 years	20.9%	20.6%	20.4%
Years in School in Current Position	4–5 years	13.6%	12.9%	12.4%

Variable	Category	% Eligible Population (*n* = 98,557)	% Who Provided Valid Responses to the Survey Weighted by Adjusted Weights (*n* = 2,588)	% Who Provided Valid Responses to the Survey Weighted by Sampling Weights (*n* = 2,588)
Years in School in Current Position	6 or more years	47.8%	45.2%	47.7%
Grades PK–2	No	50.3%	49.3%	51.9%
Grades PK–2	Yes	49.7%	50.7%	48.1%
Grades 3–5	No	48.7%	47.6%	50.4%
Grades 3–5	Yes	51.3%	52.4%	49.6%
Grades 6–8	No	38.5%	35.3%	38.2%
Grades 6–8	Yes	61.5%	64.7%	61.8%
Grades 9–12	No	68.6%	75.7%	72.6%
Grades 9–12	Yes	31.4%	24.3%	27.4%

Note. The statistics for the estimated eligible population were treated as a fixed value in the statistical testing. The statistics that were statistically significant from the corresponding value from the full sample are in **bold**. Statistics are not shown (–) because the group being analyzed is too small. ELA = English language arts.

The final adjusted weights were applied in the survey analysis so that the weighted results accurately represent the target population. Descriptive statistics were primarily employed—tabulations and cross tabulations—to answer the research questions. Consider a table with *R* rows and *C* columns, where each cell is identified by a row index (*r*) and column index (*c*), and we have data for *m* individuals in the sample. We define an indicator variable $y_{(rc)j}$ = 1 if the jth individual belongs to the cell (*r, c*), and 0 if otherwise. Weighted cell count $\widehat{N}_{(rc)j}$ is estimated as

$$\widehat{N}_{rc} = \sum_{j=1}^{m} FAW_j y_{(rc)j}$$

where FAW_j is the final adjusted weight as described in the previous section. And the weighted proportion of cell (*r, c*) is estimated as

$$\hat{p}_{rc} = \frac{\widehat{N}_{rc}}{\sum_{r=1}^{R}\sum_{c=1}^{C}\widehat{N}_{rc}}$$

which is a ratio estimator.

Appendix F. Survey Results

To analyze survey data by educator and school characteristics, educators' survey responses about their districts' educator evaluation systems were cross tabulated with educator demographic and professional characteristics, as well as the school level, locale, and student composition. This appendix presents tables of those cross tabulations.

Exhibit F–1. Percentage of Educators Reporting Evaluation System Components, by Educator Race/Ethnicity

Which *components* did the district's evaluation system use to rate your professional practice during the previous school year (2022–23)? Please select all that apply:	Educator Race/Ethnicity							
	All	Black or African American	American Indian or Alaska Native	Asian	Hispanic or Latinx	Native Hawaiian or Pacific Islander	White	Other
Standardized statewide assessments	28.8% (1.7)	26.4% (4.0)	–	27.7% (7.6)	36.2% (5.8)	–	28.2% (1.8)	33.3% (9.4)
Districtwide or locally developed student assessments	54.0% (2.2)	36.2% (3.9)	–	33.9% (12.6)	47.2% (7.4)	–	56.9% (2.3)	51.3% (9.6)
Observation rubrics for professional practice	76.6% (1.3)	76.8% (2.9)	–	77.1% (8.6)	75.4% (5.6)	–	76.5% (1.5)	81.9% (7.3)
Other measures of educator practice	8.4% (0.8)	7.7% (4.1)	–	14.1% (6.3)	4.7% (3.1)	–	8.8% (0.8)	3.4% (1.9)

Note. Number of respondents = 2,385. Percentages are based on the weighted number of respondents in each racial and ethnic group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed was too small.

Exhibit F–2. Percentage of Educators Reporting Evaluation System Components, by Educator Role

Which *components* did the district's evaluation system use to rate your professional practice during the previous school year (2022–23)? Please select all that apply:	Educator Role				
	All	Nontenured Teacher	Tenured Teacher	School Administrator	Related Service Provider
Standardized statewide assessments	28.6% (1.8)	29.2% (2.0)	28.5% (2.6)	26.9% (5.6)	23.0% (10.3)
Districtwide or locally developed student assessments	54.2% (2.3)	50.2% (3.1)	58.4% (2.5)	50.8% (5.7)	22.9% (8.3)
Observation rubrics for professional practice	77.3% (1.3)	81.1% (1.5)	75.2% (2.0)	66.8% (5.7)	84.7% (7.0)
Other measures of educator practice	8.5% (0.8)	6.6% (1.0)	9.8% (1.3)	12.3% (3.6)	0.0% (0.0)

Note. Number of respondents = 2,134. Percentages are based on the weighted number of respondents in each role. Standard errors are presented in parentheses. "School administrator" includes respondents who selected the roles of principal, assistant principal, and other school-based administrator. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–3. Percentage of Educators Reporting Evaluation System Components, by Union Representation

Which *components* did the district's evaluation system use to rate your professional practice during the previous school year (2022–23)? Please select all that apply:	Union Representation in 2022–23					
	All	Illinois Education Association	Illinois Federation of Teachers	Chicago Teachers Union	Other	Unaffiliated
Standardized statewide assessments	29.0% (1.8)	31.9% (2.3)	29.8% (3.0)	24.9% (6.3)	19.9% (4.3)	13.4% (4.2)
Districtwide or locally developed student assessments	55.8% (2.5)	60.2% (2.9)	58.9% (3.4)	40.8% (7.2)	59.5% (8.2)	48.3% (15.6)
Observation rubrics for professional practice	78.3% (1.4)	77.3% (1.7)	78.6% (2.4)	85.1% (4.2)	70.1% (6.1)	79.9% (7.5)
Other measures of educator practice	8.6% (1.0)	8.3% (1.0)	8.3% (1.9)	9.6% (2.8)	20.8% (4.7)	6.1% (3.4)

Note. Number of respondents = 1,916. Percentages are based on the weighted number of respondents in each union group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–4. Percentage of Educators Reporting Evaluation System Components, by School Level

Which *components* did the district's evaluation system use to rate your professional practice during the previous school year (2022–23)? Please select all that apply:	School Level			
	All	Elementary	Middle	High
Standardized statewide assessments	28.8% (1.7)	33.3% (1.9)	31.1% (3.2)	16.2% (2.3)
Districtwide or locally developed student assessments	54.0% (2.2)	50.0% (3.0)	63.1% (3.5)	56.1% (2.7)
Observation rubrics for professional practice	76.6% (1.3)	75.7% (1.6)	78.1% (2.5)	77.5% (2.2)
Other measures of educator practice	8.4% (0.8)	9.9% (1.1)	6.9% (2.3)	6.0% (0.9)

Note. Number of respondents = 2,385. Percentages based on the weighted number of respondents in each tested school-level group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–5. Percentage of Educators Reporting Confidence in Understanding Evaluation System Components, by Educator Race/Ethnicity

How *confident* did you feel in your understanding of the district's evaluation system during the previous school year (2022–23)?	Educator Race/Ethnicity							
	All	Black or African American	American Indian or Alaska Native	Asian	Hispanic or Latinx	Native Hawaiian or Pacific Islander	White	Other
Not confident	8.7% (0.9)	22.5% (6.7)	–	–	10.2% (3.9)	–	7.2% (0.7)	–
Somewhat confident	39.2% (1.5)	37.0% (5.2)	–	57.7% (5.9)	46.5% (6.5)	–	37.9% (1.7)	48.1% (9.9)
Extremely confident	52.1% (1.5)	40.5% (3.5)	–	41.0% (5.7)	43.3% (6.0)	–	54.9% (1.6)	34.8% (9.4)

Note. Number of respondents = 2,434. Percentages are based on the weighted number of respondents in each ethnic group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–6. Percentage of Educators Reporting Confidence in Understanding Evaluation System Components, by Educator Role

How *confident* did you feel in your understanding of the district's evaluation system during the previous school year (2022–23)?	Educator Role				
	All	Nontenured Teacher	Tenured Teacher	School Administrator	Related Service Provider
Not confident	8.9% (0.9)	9.5% (1.7)	8.1% (0.9)	–	–
Somewhat confident	38.7% (1.6)	44.1% (2.4)	35.0% (2.1)	–	–
Extremely confident	52.4% (1.6)	46.4% (2.3)	56.8% (2.2)	56.7% (5.7)	43.2% (9.7)

Note. Number of respondents = 2,124. Percentages are based on the weighted number of respondents in each role. Standard errors are presented in parentheses. "School administrator" includes respondents who selected the roles of principal, assistant principal, and other school-based administrator. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–7. Percentage of Educators Reporting Confidence in Understanding Evaluation System Components, by Chicago Public Schools

How *confident* did you feel in your understanding of the district's evaluation system during the previous school year (2022–23)?	School in Chicago Public Schools		
	All	Yes	No
Not confident	8.7% (0.9)	16.4% (4.2)	7.2% (0.6)
Somewhat confident	39.2% (1.5)	42.4% (4.8)	38.5% (1.6)
Extremely confident	52.1% (1.5)	41.2% (4.5)	54.2% (1.7)

Note. Number of respondents = 2,434. Percentages are based on the weighted number of respondents in Chicago Public Schools (CPS) and non-CPS school status. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–8. Percentage of Educators Reporting Frequency of Observation, by Educator Race/Ethnicity

How *frequently* was your professional practice *observed* as part of your district's evaluation system during the previous school year (2022–23)? Please select one:	Educator Race/Ethnicity							
	All	Black or African American	American Indian or Alaska Native	Asian	Hispanic or Latinx	Native Hawaiian or Pacific Islander	White	Other
Never	2.4% (0.5)	-	–	–	–	–	2.3% (0.5)	–
Once	18.2% (1.7)	6.3% (2.7)	–	24.1% (10.2)	8.2% (2.6)	–	20.4% (1.7)	–
Two or more times	79.4% (1.9)	90.0% (3.4)	–	74.8% (10.2)	90.9% (2.7)	–	77.4% (1.9)	86.5% (6.4)

Note. Number of respondents = 2,535. Percentages are based on the weighted number of respondents in each ethnic group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–9. Percentage of Educators Reporting Frequency of Observation, by Educator Role

How *frequently* was your professional practice *observed* as part of your district's evaluation system during the previous school year (2022–23)? Please select one:	Educator Role				
	All	Nontenured Teacher	Tenured Teacher	School Administrator	Related Service Provider
Never	1.7% (0.4)	–	2.1% (0.5)	6.1% (2.9)	–
Once	18.0% (1.8)	–	30.5% (3.5)	16.8% (4.7)	–
Two or more times	80.2% (1.9)	97.5% (0.9)	67.4% (3.6)	77.0% (4.6)	–

Note. Number of respondents = 2,104. Percentages are based on the weighted number of respondents in each role. Standard errors are presented in parentheses. "School administrator" includes respondents who selected the roles of principal, assistant principal, and other school-based administrator. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–10. Percentage of Educators Reporting Frequency of Observation, by Educator Years of Experience

How *frequently* was your professional practice *observed* as part of your district's evaluation system during the previous school year (2022–23)? Please select one:	Educator Years of Experience at End of 2022–23				
	All	First Year	Second or Third Year	Fourth or Fifth Year	Sixth or More Year
Never	2.4% (0.5)	–	–	–	3.0% (0.6)
Once	18.2% (1.7)	–	–	–	31.5% (3.2)
Two or more times	79.4% (1.9)	90.7% (2.0)	93.7% (1.2)	86.9% (2.4)	65.4% (3.5)

Note. Number of respondents = 2,535. Percentages are based on the weighted number of respondents in each years-of-experience group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–11. Percentage of Educators Reporting Frequency of Observation, by School Service Area

How *frequently* was your professional practice *observed* as part of your district's evaluation system during the previous school year (2022–23)? Please select one:	School Service Area						
	All	1	2	3	4	5	6
Never	2.4% (0.5)	2.3% (0.7)	–	–	–	–	–
Once	18.2% (1.7)	18.5% (2.3)	–	–	–	–	–
Two or more times	79.4% (1.9)	79.2% (2.5)	69.0% (3.3)	87.1% (6.3)	84.8% (3.8)	86.8% (3.1)	76.0% (5.0)

Note. Number of respondents = 2,535. Percentages are based on the weighted number of respondents in each service area. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–12. Percentage of Educators Reporting Frequency of Preconference, by Educator Role

How *frequently* did you engage in a *preconference before* you were observed as part of your district's evaluation system during the previous school year (2022–23)? Please select one:	Educator Role				
	All	Nontenured Teacher	Tenured Teacher	School Administrator	Related Service Provider
Never	7.8% (0.8)	2.4% (0.8)	9.9% (1.3)	28.4% (4.8)	–
Once	41.4% (1.6)	11.2% (1.3)	64.8% (2.7)	37.0% (5.0)	–
Two or more times	50.9% (1.6)	86.4% (1.7)	25.2% (2.4)	34.6% (5.4)	52.0% (12.1)

Note. Number of respondents = 2,116. Percentages are based on the weighted number of respondents in each role. Standard errors are presented in parentheses. "School administrator" includes respondents who selected the roles of principal, assistant principal, and other school-based administrator. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–13. Percentage of Educators Reporting Frequency of Preconference, by Educator Years of Experience

How *frequently* did you engage in a *preconference before* you were observed as part of your district's evaluation system during the previous school year (2022–23)? Please select one:	Educator Years of Experience at End of 2022–23				
	All	First Year	Second or Third Year	Fourth or Fifth Year	Sixth or More Year
Never	8.2% (0.9)	6.6% (1.7)	4.5% (1.2)	8.7% (1.7)	10.5% (1.4)
Once	41.4% (1.6)	12.9% (2.0)	18.2% (2.2)	40.4% (3.7)	65.4% (2.5)
Two or more times	50.4% (1.5)	80.5% (2.5)	77.3% (2.3)	50.9% (3.7)	24.0% (2.1)

Note. Number of respondents = 2,558. Percentages are based on the weighted number of respondents in each years-of-experience group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–14. Percentage of Educators Reporting Frequency of Preconference, by Union Representation

How *frequently* did you engage in a *preconference before* you were observed as part of your district's evaluation system during the previous school year (2022–23)? Please select one:	Union Representation in 2022–23					
	All	Illinois Education Association	Illinois Federation of Teachers	Chicago Teachers Union	Other	Unaffiliated
Never	7.9% (0.9)	8.1% (1.3)	7.5% (1.3)		–	–
Once	44.2% (1.7)	40.0% (1.9)	50.5% (3.0)		–	–
Two or more times	47.9% (1.7)	51.9% (2.1)	42.0% (2.5)	43.2% (6.0)	61.1% (6.8)	21.3% (8.3)

Note. Number of respondents = 2,558. Percentages are based on the weighted number of respondents in each union group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–15. Percentage of Educators Reporting Frequency of Preconference, by Chicago Public Schools

How *frequently* did you engage in a *preconference before* you were observed as part of your district's evaluation system during the previous school year (2022–23)? Please select one:	School in Chicago Public Schools		
	All	Yes	No
Never	–	7.3% (3.5)	8.4% (0.8)
Once	41.4% (1.6)	52.5% (5.2)	39.2% (1.4)
Two or more times	–	40.2% (6.3)	52.4% (1.4)

Note. Number of respondents = 2,558. Percentages are based on the weighted number of respondents in Chicago Public Schools (CPS) and non-CPS school status. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–16. Percentage of Educators Reporting Frequency of Postconference, by Educator Role

How *frequently* did you engage in a *postconference after* you were observed as part of your district's evaluation system during the previous school year (2022–23)? Please select one:	Educator Role				
	All	Nontenured Teacher	Tenured Teacher	School Administrator	Related Service Provider
Never	6.3% (0.9)	–	7.8% (1.2)	19.0% (4.4)	–
Once	40.2% (1.8)	–	62.8% (3.0)	36.3% (5.0)	–
Two or more times	53.5% (1.8)	86.1% (1.7)	29.4% (2.9)	44.7% (5.3)	57.6% (12.1)

Note. Number of respondents = 2,118. Percentages are based on the weighted number of respondents in each role. "School administrator" includes respondents who selected the roles of principal, assistant principal, and other school-based administrator. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–17. Percentage of Educators Reporting Frequency of Postconference, by Educator Years of Experience

How *frequently* did you engage in a *postconference after* you were observed as part of your district's evaluation system during the previous school year (2022–23)? Please select one:	Educator Years of Experience at End of 2022–23				
	All	First Year	Second or Third Year	Fourth or Fifth Year	Sixth or More Year
Never	6.5% (0.9)	5.1% (1.6)	4.7% (1.5)	6.5% (1.4)	8.0% (1.3)
Once	40.6% (1.8)	12.3% (1.9)	18.7% (2.3)	38.8% (4.7)	64.3% (2.8)
Two or more times	52.9% (1.8)	82.6% (2.4)	76.6% (2.5)	54.7% (4.3)	27.7% (2.6)

Note. Number of respondents = 2,554. Percentages are based on the weighted number of respondents in each years-of-experience group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–18. Percentage of Educators Reporting Frequency of Postconference, by Union Representation

How *frequently* did you engage in a *postconference after* you were observed as part of your district's evaluation system during the previous school year (2022–23)? Please select one:	Union Representation in 2022–23					
	All	Illinois Education Association	Illinois Federation of Teachers	Chicago Teachers Union	Other	Unaffiliated
Never	6.4% (1.0)	5.9% (1.1)	6.6% (1.6)		9.2% (4.1)	–
Once	42.9% (1.9)	39.2% (1.9)	48.1% (2.7)		23.1% (5.9)	–
Two or more times	50.7% (2.0)	54.9% (2.2)	45.3% (2.1)	43.0% (8.1)	67.7% (6.7)	38.0% (7.9)

Note. Number of respondents = 1,903. Percentages are based on the weighted number of respondents in each union group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–19. Percentage of Educators Reporting Frequency of Written Feedback, by Educator Role

How *frequently* did you receive *written feedback* based on an observation as part of your district's evaluation system during the previous school year (2022–23)? Please select one:	Educator Role				
	All	Nontenured Teacher	Tenured Teacher	School Administrator	Related Service Provider
Never	5.2% (0.9)	3.1% (1.0)	6.7% (1.3)	5.7% (1.8)	–
Once	30.0% (2.3)	10.1% (1.7)	45.2% (4.0)	29.3% (5.1)	–
Two or more times	64.8% (2.7)	86.8% (2.2)	48.2% (4.4)	65.0% (5.3)	55.9% (12.1)

Note. Number of respondents = 2,108. Percentages are based on the weighted number of respondents in each role. Standard errors are presented in parentheses. "School administrator" includes respondents who selected the roles of principal, assistant principal, and other school-based administrator. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–20. Percentage of Educators Reporting Frequency of Written Feedback, by Educator Years of Experience

How *frequently* did you receive *written feedback* based on an observation as part of your district's evaluation system during the previous school year (2022–23)? Please select one:	Educator Years of Experience at End of 2022–23				
	All	First Year	Second or Third Year	Fourth or Fifth Year	Sixth or More Year
Never	5.6% (0.9)	4.4% (1.6)	2.4% (0.9)	6.5% (2.1)	7.4% (1.3)
Once	30.4% (2.1)	11.2% (2.3)	17.1% (2.3)	27.0% (3.2)	46.4% (3.8)
Two or more times	64.0% (2.5)	84.4% (2.7)	80.4% (2.5)	66.5% (3.1)	46.1% (4.1)

Note. Number of respondents = 2,550. Percentages are based on the weighted number of respondents in each years-of-experience group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–21. Percentage of Educators Reporting Frequency of Written Feedback, by Chicago Public Schools

How *frequently* did you receive written *feedback* based on an observation as part of your district's evaluation system during the previous school year (2022–23)? Please select one:	School in Chicago Public Schools		
	All	Yes	No
Never	5.6% (0.9)	–	4.9% (0.8)
Once	30.4% (2.1)	–	27.9% (2.1)
Two or more times	64.0% (2.5)	47.4% (7.4)	67.2% (2.5)

Note. Number of respondents = 2,550. Percentages are based on the weighted number of respondents in Chicago Public Schools (CPS) and non-CPS school status. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–22. Percentage of Educators Reporting Frequency of Being Offered Coaching, by Educator Role

How frequently were you *offered coaching* as a result of an observation as part of your district's evaluation system during the previous school year (2022–23)? Please select one:	Educator Role				
	All	Nontenured Teacher	Tenured Teacher	School Administrator	Related Service Provider
Never	70.1% (1.7)	56.6% (2.8)	79.9% (1.5)	70.6% (5.3)	64.3% (10.8)
Once	12.5% (0.9)	14.0% (1.8)	11.6% (1.3)	10.2% (2.8)	–
Two or more times	17.4% (1.5)	29.3% (2.8)	8.5% (1.0)	19.2% (4.7)	–

Note. Number of respondents = 1,944. Percentages are based on the weighted number of respondents in each role. Standard errors are presented in parentheses. "School administrator" includes respondents who selected the roles of principal, assistant principal, and other school-based administrator. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–23. Percentage of Educators Reporting Frequency of Being Offered Coaching, by Educator Years of Experience

How frequently were you *offered coaching* as a result of an observation as part of your district's evaluation system during the previous school year (2022–23)? Please select one:	Educator Years of Experience at End of 2022–23				
	All	First Year	Second or Third Year	Fourth or Fifth Year	Sixth or More Year
Never	69.9% (1.7)	50.5% (3.4)	63.9% (3.2)	71.8% (3.7)	80.8% (1.4)
Once	12.5% (1.0)	15.1% (2.3)	14.2% (2.1)	13.5% (3.0)	10.4% (1.2)
Two or more times	17.5% (1.5)	34.4% (3.2)	21.9% (2.7)	14.7% (2.7)	8.8% (1.1)

Note. Number of respondents = 2,258. Percentages are based on the weighted number of respondents in each years-of-experience group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–24. Percentage of Educators Reporting Actionability of Written and Verbal Postconference Feedback, by Educator Role

How actionable was written and verbal feedback?	Educator Role				
	All	Nontenured Teacher	Tenured Teacher	School Administrator	Related Service Provider
Not at all actionable	10.8% (1.2)	5.9% (1.4)	15.3% (1.5)	–	–
Somewhat actionable	55.9% (1.6)	55.8% (2.7)	56.4% (1.9)	–	–
Extremely actionable	33.4% (1.6)	38.3% (2.3)	28.3% (2.0)	38.3% (6.4)	62.7% (11.5)

Note. Number of respondents = 1,919. Percentages are based on the weighted number of respondents in each role. Standard errors are presented in parentheses. "School administrator" includes respondents who selected the roles of principal, assistant principal, and other school-based administrator. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–25. Percentage of Educators Reporting Actionability of Written and Verbal Postconference Feedback, by School Service Area

How actionable was written and verbal feedback?	School Service Area						
	All	1	2	3	4	5	6
Not at all actionable	10.6% (1.1)	10.4% (1.4)	13.3% (4.3)	12.3% (3.4)	10.9% (2.1)	9.1% (1.3)	–
Somewhat actionable	55.9% (1.6)	57.3% (1.9)	52.1% (3.2)	60.4% (6.1)	56.9% (3.9)	52.2% (6.9)	–
Extremely actionable	33.4% (1.5)	32.3% (1.9)	34.5% (3.8)	27.3% (5.7)	32.3% (4.1)	38.7% (6.2)	56.6% (3.8)

Note. Number of respondents = 2,213. Percentages are based on the weighted number of respondents in each service area. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–26. Percentage of Educators Reporting Actionability of Written and Verbal Postconference Feedback, by School Urbanicity

How actionable was written and verbal feedback?	School Urbanicity				
	All	City	Suburb	Town	Rural
Not at all actionable	10.6% (1.1)	13.0% (3.4)	10.4% (1.3)	7.3% (1.5)	10.0% (1.8)
Somewhat actionable	55.9% (1.6)	62.9% (4.6)	54.3% (2.0)	51.8% (4.3)	54.2% (2.7)
Extremely actionable	33.4% (1.5)	24.0% (2.7)	35.3% (2.2)	41.0% (4.2)	35.7% (2.9)

Note. Number of respondents = 2,213. Percentages are based on the weighted number of respondents in schools in each urbanicity category. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–27. Percentage of Educators Reporting Actionability of Written and Verbal Postconference Feedback, by Chicago Public Schools

How actionable was written and verbal feedback?	School in Chicago Public Schools		
	All	Yes	No
Not at all actionable	10.6% (1.1)	13.3% (4.8)	10.1% (0.9)
Somewhat actionable	55.9% (1.6)	67.3% (6.6)	53.8% (1.5)
Extremely actionable	33.4% (1.5)	19.4% (3.6)	36.1% (1.6)

Note. Number of respondents = 2,213. Percentages are based on the weighted number of respondents in Chicago Public Schools (CPS) and non-CPS school status. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–28. Percentage of Educators Reporting Usefulness of Coaching, by Educator Years of Experience

How useful was the coaching?	Educator Years of Experience at End of 2022–23				
	All	First Year	Second or Third Year	Fourth or Fifth Year	Sixth or More Year
Not at all useful	11.5% (2.0)	8.7% (3.1)	–	–	16.1% (3.8)
Somewhat useful	50.1% (2.8)	41.8% (5.1)	–	–	55.8% (4.3)
Extremely useful	38.5% (2.3)	49.4% (4.4)	42.0% (5.6)	25.8% (5.2)	28.0% (4.4)

Note. Number of respondents = 643. Percentages are based on the weighted number of respondents in each years-of-experience group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–29. Percentage of Educators Reporting Supportiveness of District's Evaluation System, by Educator Role

How supportive was the evaluation system?	Educator Role				
	All	Nontenured Teacher	Tenured Teacher	School Administrator	Related Service Provider
Not at all supportive	20.4% (1.6)	14.5% (1.8)	25.1% (2.0)	18.5% (5.0)	–
Somewhat supportive	49.2% (1.4)	48.2% (2.1)	49.0% (1.9)	57.5% (4.9)	58.0% (12.1)
Extremely supportive	30.4% (1.5)	37.3% (1.9)	25.8% (1.8)	24.0% (4.5)	–

Notes. Number of respondents = 2,094. Percentages are based on the weighted number of respondents in each role. Standard errors are presented in parentheses. "School administrator" includes respondents who selected the roles of principal, assistant principal, and other school-based administrator. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–30. Percentage of Educators Reporting Supportiveness of District's Evaluation System, by Chicago Public Schools

How supportive was the evaluation system?	School in Chicago Public Schools		
	All	Yes	No
Not at all supportive	20.2% (1.6)	34.3% (7.3)	17.6% (1.4)
Somewhat supportive	49.0% (1.3)	46.3% (5.3)	49.5% (1.1)
Extremely supportive	30.8% (1.5)	19.4% (4.1)	32.9% (1.6)

Note. Number of respondents = 2,400. Percentages are based on the weighted number of respondents in Chicago Public Schools (CPS) and non-CPS school status. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–31. Percentage of Educators Reporting Supportiveness of District's Evaluation System, by School English Language Arts Proficiency Rate Quartile: 2022–23

How supportive was the evaluation system?	School English Language Arts Proficiency Rate Quartile: 2022–23				
	All	First (Percentiles 1–25)	Second (Percentiles 26–50)	Third (Percentiles 51–75)	Fourth (Percentiles 76–99)
Not at all supportive	20.1% (1.7)	27.5% (3.0)	13.1% (1.9)	15.4% (2.0)	20.6% (3.7)
Somewhat supportive	49.0% (1.3)	46.9% (2.5)	52.5% (2.5)	52.7% (2.3)	44.4% (3.0)
Extremely supportive	30.9% (1.5)	25.6% (2.6)	34.4% (2.8)	31.9% (2.0)	35.0% (3.4)

Note. Number of respondents = 2,286. Percentages are based on the weighted number of respondents in schools in each English language arts proficiency rate quartile. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–32. Percentage of Educators Reporting Fairness of Student Assessment Component, by Subject Area

How *fairly* did the *student assessment component* of the district's evaluation system measure your professional practice during the previous school year (2022–23)? Please select one:	Subject Area Taught in 2022–23				
	All	General Education	Special Education	Core Subject	Elective Subject
Not fairly at all	16.7% (1.6)	17.7% (3.2)	23.0% (4.0)	15.7% (2.5)	15.6% (3.2)
Somewhat fairly	48.3% (2.1)	54.3% (4.3)	40.7% (3.7)	48.3% (2.5)	41.7% (3.0)
Extremely fairly	35.0% (2.2)	28.0% (3.0)	36.3% (3.3)	36.1% (2.8)	42.8% (3.6)

Note. Number of respondents = 1,485. Percentages are based on the weighted number of respondents in each role. Standard errors are presented in parentheses. "Core subject" includes English language arts, mathematics, science, and history/social science; "Elective subject" includes visual/performing arts, physical/health education, foreign language, career and technical education, and other. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–33. Percentage of Educators Reporting Fairness of Student Assessment Component, by School Service Area

How *fairly* did the *student assessment component* of the district's evaluation system measure your professional practice during the previous school year (2022–23)? Please select one:	School Service Area						
	All	1	2	3	4	5	6
Not fairly at all	16.3% (1.5)	19.1% (2.0)	11.6% (3.9)	–	9.1% (2.6)	13.2% (2.2)	–
Somewhat fairly	48.8% (2.0)	49.2% (2.8)	47.8% (3.3)	–	52.2% (5.4)	36.9% (4.3)	–
Extremely fairly	34.9% (2.1)	31.7% (2.8)	40.6% (4.7)	35.5% (5.4)	38.7% (4.9)	49.8% (4.3)	36.5% (9.0)

Note. Number of respondents = 1,485. Percentages are based on the weighted number of respondents in each service area. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–34. Percentage of Educators Reporting Fairness of Student Assessment Component, by School Urbanicity

How *fairly* did the *student assessment component* of the district's evaluation system measure your professional practice during the previous school year (2022–23)? Please select one:	School Urbanicity				
	All	City	Suburb	Town	Rural
Not fairly at all	16.3% (1.5)	25.5% (4.1)	15.2% (1.8)	13.2% (3.1)	9.9% (2.6)
Somewhat fairly	48.8% (2.0)	52.0% (8.2)	48.6% (1.8)	46.5% (5.0)	46.3% (3.1)
Extremely fairly	34.9% (2.1)	22.5% (6.5)	36.2% (2.7)	40.3% (4.7)	43.7% (4.0)

Note. Number of respondents = 1,485. Percentages are based on the weighted number of respondents in schools in each urbanicity category. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–35. Percentage of Educators Reporting Fairness of Student Assessment Component, by Chicago Public Schools

How *fairly* did the *student assessment component* of the district's evaluation system measure your professional practice during the previous school year (2022–23)? Please select one:	School in Chicago Public Schools		
	All	Yes	No
Not fairly at all	16.3% (1.5)	28.4% (5.5)	14.4% (1.4)
Somewhat fairly	48.8% (2.0)	48.7% (11.5)	48.8% (1.4)
Extremely fairly	34.9% (2.1)	22.9% (9.0)	36.9% (2.0)

Note. Number of respondents = 1,485. Percentages are based on the weighted number of respondents in Chicago Public Schools (CPS) and non-CPS school status. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–36. Percentage of Educators Reporting Fairness of Student Assessment Component, by School English Language Arts Proficiency Rate Quartile: 2022–23

How *fairly* did the *student assessment component* of the district's evaluation system measure your professional practice during the previous school year (2022–23)? Please select one:	School English Language Arts Proficiency Rate Quartile: 2022–23				
	All	First (Percentiles 1–25)	Second (Percentiles 26–50)	Third (Percentiles 51–75)	Fourth (Percentiles 76–99)
Not fairly at all	16.6% (1.6)	20.2% (1.9)	12.5% (2.9)	18.9% (4.1)	11.4% (2.2)
Somewhat fairly	48.3% (2.1)	50.0% (3.7)	52.5% (3.3)	47.8% (5.1)	42.5% (2.7)
Extremely fairly	35.1% (2.2)	29.8% (4.1)	35.0% (3.3)	33.3% (4.4)	46.1% (3.3)

Note. Number of respondents = 1,425. Percentages are based on the weighted number of respondents in schools in each English language arts proficiency rate quartile. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–37. Percentage of Educators Reporting Fairness of Student Assessment Component, by School Mathematics Proficiency Rate Quartile: 2022–23

How *fairly* did the *student assessment component* of the district's evaluation system measure your professional practice during the previous school year (2022–23)? Please select one:	School Mathematics Proficiency Rate Quartile: 2022–23				
	All	First (Percentiles 1–25)	Second (Percentiles 26–50)	Third (Percentiles 51–75)	Fourth (Percentiles 76–99)
Not fairly at all	16.6% (1.6)	20.4% (1.9)	11.1% (3.4)	20.2% (4.3)	12.2% (1.9)
Somewhat fairly	48.3% (2.1)	48.6% (3.4)	55.9% (4.1)	45.6% (5.9)	44.9% (2.4)
Extremely fairly	35.1% (2.2)	30.9% (3.7)	33.0% (6.3)	34.2% (4.2)	42.9% (3.0)

Note. Number of respondents = 1,425. Percentages are based on the weighted number of respondents in schools in each mathematics proficiency rate quartile. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–38. Percentage of Educators Reporting Formal Observation Fairness, by Educator Race/Ethnicity

How *fairly* did the *formal observation component* of the district's evaluation system measure your professional practice during the previous school year (2022–23)? Please select one:	Educator Race/Ethnicity							
	All	Black or African American	American Indian or Alaska Native	Asian	Hispanic or Latinx	Native Hawaiian or Pacific Islander	White	Other
Not fairly at all	7.2% (1.0)	23.8% (6.2)	–	–	–	–	5.9% (0.9)	2.5% (2.2)
Somewhat fairly	42.0% (1.9)	45.6% (7.5)	–	–	–	–	41.0% (1.9)	43.0% (10.8)
Extremely fairly	50.8% (1.9)	30.6% (7.7)	–	60.4% (8.6)	40.4% (10.3)	–	53.1% (2.0)	54.5% (10.6)

Note. Number of respondents = 1,724. Percentages are based on the weighted number of respondents in each ethnic group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–39. Percentage of Educators Reporting Formal Observation Fairness, by Union Representation

How *fairly* did the *formal observation component* of the district's evaluation system measure your professional practice during the previous school year (2022–23)? Please select one:	Union Representation in 2022–23					
	All	Illinois Education Association	Illinois Federation of Teachers	Chicago Teachers Union	Other	Unaffiliated
Not fairly at all	7.5% (1.0)	4.8% (1.0)	5.4% (1.4)	–	–	–
Somewhat fairly	42.5% (1.9)	37.9% (1.9)	41.9% (3.1)	–	–	–
Extremely fairly	50.0% (1.9)	57.2% (2.1)	52.6% (2.9)	31.7% (6.7)	51.1% (7.6)	38.6% (9.3)

Note. Number of respondents = 1,428. Percentages are based on the weighted number of respondents in each union group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–40. Percentage of Educators Reporting Formal Observation Fairness, by School Urbanicity

How *fairly* did the *formal observation component* of the district's evaluation system measure your professional practice during the previous school year (2022–23)? Please select one:	School Urbanicity				
	All	City	Suburb	Town	Rural
Not fairly at all	7.2% (1.0)	13.0% (2.6)	5.2% (1.2)	5.7% (2.6)	6.3% (2.3)
Somewhat fairly	42.0% (1.9)	51.7% (6.4)	39.9% (1.9)	36.0% (2.9)	37.6% (4.0)
Extremely fairly	50.8% (1.9)	35.3% (5.6)	54.9% (2.1)	58.4% (3.5)	56.1% (4.7)

Note. Number of respondents = 1,724. Percentages are based on the weighted number of respondents in schools in each urbanicity category. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–41. Percentage of Educators Reporting Formal Observation Fairness, by Chicago Public Schools

How *fairly* did the *formal observation component* of the district's evaluation system measure your professional practice during the previous school year (2022–23)? Please select one:	School in Chicago Public Schools		
	All	Yes	No
Not fairly at all	7.2% (1.0)	14.1% (3.4)	5.8% (0.9)
Somewhat fairly	42.0% (1.9)	55.8% (8.5)	39.0% (1.5)
Extremely fairly	50.8% (1.9)	30.2% (7.4)	55.2% (1.6)

Note. Number of respondents = 1,724. Percentages are based on the weighted number of respondents in Chicago Public Schools (CPS) and non-CPS school status. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–42. Percentage of Educators Reporting Accuracy of District's Evaluation System, by Educator Race/Ethnicity

How *accurately* did the district's evaluation system measure the effectiveness of your overall professional practice during the previous school year (2022–23)? Please select one:	Educator Race/Ethnicity							
	All	Black or African American	American Indian or Alaska Native	Asian	Hispanic or Latinx	Native Hawaiian or Pacific Islander	White	Other
Not at all accurately	10.5% (0.9)	14.8% (4.5)	–	–	11.3% (4.4)	–	10.0% (1.0)	–
Somewhat accurately	50.0% (1.4)	60.6% (5.7)	–	–	55.2% (6.0)	–	48.6% (1.3)	–
Extremely accurately	39.5% (1.5)	24.7% (6.3)	–	44.5% (7.2)	33.5% (6.3)	–	41.4% (1.5)	32.2% (9.9)

Note. Number of respondents = 2,260. Percentages are based on the weighted number of respondents in each ethnic group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–43. Percentage of Educators Reporting Accuracy of District's Evaluation System, by Union Representation

How *accurately* did the district's evaluation system measure the effectiveness of your overall professional practice during the previous school year (2022–23)? Please select one:	Union Representation in 2022–23					
	All	Illinois Education Association	Illinois Federation of Teachers	Chicago Teachers Union	Other	Unaffiliated
Not at all accurately	11.2% (1.0)	8.1% (1.0)	9.8% (1.5)	21.6% (3.6)		–
Somewhat accurately	50.0% (1.4)	48.7% (1.6)	49.2% (2.9)	49.1% (4.8)		–
Extremely accurately	38.9% (1.5)	43.2% (1.8)	41.0% (3.3)	29.3% (5.5)	31.8% (5.2)	31.4% (6.9)

Note. Number of respondents = 1,856. Percentages are based on the weighted number of respondents in each union group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–44. Percentage of Educators Reporting Accuracy of District's Evaluation System, by School Urbanicity

How *accurately* did the district's evaluation system measure the effectiveness of your overall professional practice during the previous school year (2022–23)? Please select one:	School Urbanicity				
	All	City	Suburb	Town	Rural
Not at all accurately	10.5% (0.9)	18.8% (2.5)	8.5% (1.1)	8.1% (2.3)	6.5% (1.8)
Somewhat accurately	50.0% (1.4)	51.2% (3.3)	50.1% (1.9)	48.1% (3.2)	49.5% (2.9)
Extremely accurately	39.5% (1.5)	29.9% (4.1)	41.4% (2.2)	43.8% (3.5)	44.0% (3.6)

Note. Number of respondents = 2,260. Percentages are based on the weighted number of respondents in schools in each urbanicity category. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–45. Percentage of Educators Reporting Accuracy of District's Evaluation System, by Chicago Public Schools

How *accurately* did the district's evaluation system measure the effectiveness of your overall professional practice during the previous school year (2022–23)? Please select one:	School in Chicago Public Schools		
	All	Yes	No
Not at all accurately	10.5% (0.9)	20.3% (3.5)	8.6% (0.8)
Somewhat accurately	50.0% (1.4)	51.2% (4.3)	49.8% (1.4)
Extremely accurately	39.5% (1.5)	28.5% (5.6)	41.6% (1.6)

Note. Number of respondents = 2,260. Percentages are based on the weighted number of respondents in Chicago Public Schools (CPS) and non-CPS school status. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–46. Percentage of Educators Reporting Accuracy of District's Evaluation System, by School Low-Income Enrollment Quartile: 2022–23

System Accuracy	School Low-Income Enrollment Quartile: 2022–23				
	All	First (Percentiles 1–25	Second (Percentiles 26–50)	Third (Percentiles 51–75)	Fourth (Percentiles 76–99)
Not at all accurate	10.4% (0.9)	8.1% (1.5)	7.7% (1.8)	10.3% (2.2)	14.6% (2.3)
Somewhat accurate	50.3% (1.4)	47.6% (1.9)	50.5% (1.9)	49.7% (2.7)	53.2% (3.2)
Extremely accurate	39.3% (1.6)	44.2% (2.7)	41.8% (2.6)	40.0% (3.1)	32.2% (2.7)

Note. Number of respondents = 2,202. Percentages are based on the weighted number of respondents in schools in each low-income enrollment quartile. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–47. Percentage of Educators Reporting Perception of Evaluator Knowledge, by Union Representation

How knowledgeable was your evaluator?	Union Representation in 2022–23					
	All	Illinois Education Association	Illinois Federation of Teachers	Chicago Teachers Union	Other	Unaffiliated
Not at all knowledgeable	9.1% (1.2)	7.3% (1.0)	8.7% (1.4)	12.6% (3.8)	–	–
Somewhat knowledgeable	40.1% (1.6)	39.3% (1.7)	38.0% (2.2)	43.9% (6.9)	–	–
Extremely knowledgeable	50.8% (2.0)	53.3% (2.1)	53.4% (2.5)	43.5% (8.3)	52.8% (7.0)	28.4% (7.0)

Note. Number of respondents = 1,895. Percentages are based on the weighted number of respondents in each union group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–48. Percentage of Educators Reporting Perception of Evaluator Knowledge, by School Urbanicity

How knowledgeable was your evaluator?	School Urbanicity				
	All	City	Suburb	Town	Rural
Not at all knowledgeable	8.7% (1.0)	12.5% (3.3)	8.8% (1.1)	3.7% (0.7)	6.2% (1.8)
Somewhat knowledgeable	40.4% (1.4)	43.7% (4.3)	41.3% (1.7)	38.0% (2.9)	33.0% (2.9)
Extremely knowledgeable	50.8% (1.8)	43.8% (5.2)	49.9% (2.3)	58.3% (2.8)	60.8% (2.9)

Note. Number of respondents = 2,307. Percentages are based on the weighted number of respondents in schools in each urbanicity category. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–49. Percentage of Educators Reporting Hours of Professional Learning, by Educator Gender

Think of the hours of professional learning you accessed or received during the previous school year (2022–23). How many of these hours, if any, were directly connected to feedback you received as part of your evaluation? Please select one:	Educator Gender		
	All	Male	Female
None	33.8% (1.8)	26.7% (2.4)	35.9% (2.0)
Less than 1 hour	18.8% (1.3)	17.9% (2.4)	19.0% (1.6)
1–2 hours	21.2% (1.2)	24.4% (2.1)	20.3% (1.5)
3–5 hours	12.5% (1.0)	13.3% (1.9)	12.3% (1.0)
6–10 hours	6.8% (0.7)	7.9% (1.7)	6.5% (0.8)
11 or more hours	6.9% (0.8)	9.8% (1.8)	6.0% (0.8)

Note. Number of respondents = 2,058. Percentages are based on the weighted number of respondents in each gender group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–50. Percentage of Educators Reporting Hours of Professional Learning, by Educator Years of Experience

Think of the hours of professional learning you accessed or received during the previous school year (2022–23). How many of these hours, if any, were directly connected to feedback you received as part of your evaluation? Please select one:	Educator Years of Experience at End of 2022–23				
	All	First Year	Second or Third Year	Fourth or Fifth Year	Sixth or More Year
None	33.8% (1.8)	31.7% (3.3)	29.9% (2.6)	33.4% (4.2)	36.8% (2.4)
Less than 1 hour	18.8% (1.3)	14.7% (2.5)	19.1% (2.0)	15.9% (2.6)	21.3% (1.9)
1–2 hours	21.2% (1.2)	20.4% (2.9)	22.5% (3.1)	25.3% (4.4)	19.8% (1.4)
3–5 hours	12.5% (1.0)	14.0% (2.1)	12.4% (1.7)	13.0% (2.4)	11.7% (1.3)
6–10 hours	6.8% (0.7)	6.5% (1.6)	9.1% (1.6)	6.4% (2.0)	6.0% (1.0)
11 or more hours	6.9% (0.8)	12.6% (2.2)	7.1% (2.1)	6.0% (1.7)	4.5% (1.0)

Note. Number of respondents = 2,058. Percentages are based on the weighted number of respondents in each years-of-experience group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–51. Percentage of Educators Reporting Difficulty of Getting Support, by Union Representation

If you requested *extra support* (e.g., informal observations, informal feedback, or informal mentorship from colleagues) to improve your practice, how difficult was it to get the support that you needed during the previous school year (2022–23)? Please select one:	Union Representation in 2022–23					
	All	Illinois Education Association	Illinois Federation of Teachers	Chicago Teachers Union	Other	Unaffiliated
Not at all difficult	55.8% (2.9)	61.3% (2.9)	61.4% (4.4)	32.0% (5.9)	69.6% (7.4)	–
Somewhat difficult	29.3% (2.3)	28.3% (2.7)	27.2% (3.5)	40.9% (6.9)	–	–
Extremely difficult	14.8% (1.9)	10.4% (1.8)	11.5% (2.7)	27.1% (4.2)	–	–

Note. Number of respondents = 763. Percentages are based on the weighted number of respondents in each union group. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–52. Percentage of Educators Reporting Difficulty of Getting Support, by School Urbanicity

If you requested *extra support* (e.g., informal observations, informal feedback, or informal mentorship from colleagues) to improve your practice, how *difficult* was it to get the support that you needed during the previous school year (2022–23)? Please select one:	School Urbanicity				
	All	City	Suburb	Town	Rural
Not at all difficult	57.9% (2.6)	37.2% (5.0)	64.4% (3.0)	59.8% (4.6)	71.0% (4.0)
Somewhat difficult	27.7% (2.1)	35.0% (4.7)	24.0% (2.4)	31.2% (5.3)	–
Extremely difficult	14.4% (1.5)	27.8% (3.2)	11.6% (1.7)	9.0% (2.2)	–

Note. Number of respondents = 944. Percentages are based on the weighted number of respondents in schools in each urbanicity category. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded. Statistics are not shown (–) because the group being analyzed is too small.

Exhibit F–53. Percentage of Educators Reporting Difficulty of Getting Support, by Chicago Public Schools

If you requested *extra support* (e.g., informal observations, informal feedback, or informal mentorship from colleagues) to improve your practice, how *difficult* was it to get the support that you needed during the previous school year (2022–23)? Please select one:	School in Chicago Public Schools		
	All	Yes	No
Not at all difficult	57.9% (2.6)	29.1% (5.8)	64.7% (2.1)
Somewhat difficult	27.7% (2.1)	38.3% (6.0)	25.2% (1.9)
Extremely difficult	14.4% (1.5)	32.5% (3.9)	10.1% (1.3)

Note. Number of respondents = 944. Percentages are based on the weighted number of respondents in Chicago Public Schools (CPS) and non-CPS school status. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

Exhibit F–54. Percentage of Educators Reporting Difficulty of Getting Support, by School English Language Arts Proficiency Rate Quartile: 2022–23

If you requested *extra support* (e.g., informal observations, informal feedback, or informal mentorship from colleagues) to improve your practice, how *difficult* was it to get the support that you needed during the previous school year (2022–23)? Please select one:	School English Language Arts Proficiency Rate Quartile: 2022–23				
	All	First (Percentiles 1–25)	Second (Percentiles 26–50)	Third (Percentiles 51–75)	Fourth (Percentiles 76–99)
Not at all difficult	57.7% (2.8)	48.3% (5.6)	62.2% (5.4)	63.3% (5.0)	64.5% (5.7)
Somewhat difficult	27.6% (2.1)	29.3% (3.9)	25.1% (4.1)	27.8% (3.6)	26.3% (5.2)
Extremely difficult	14.7% (1.6)	22.4% (3.0)	12.7% (3.9)	8.9% (3.2)	9.2% (2.5)

Note. Number of respondents = 903. Percentages are based on the weighted number of respondents in schools in each English language arts proficiency rate quartile. Standard errors are presented in parentheses. Respondents who did not select a response or selected "unsure" were excluded.

References

The American Association for Public Opinion Research. (2023). *Standard definitions: Final dispositions of case codes and outcome rates for surveys.* (*10th Edition*.) AAPOR. https://aapor.org/wp-content/uploads/2023/05/Standards-Definitions-10th-edition.pdf

Cochran, W. G. (1977). *Sampling techniques*. John Wiley & Sons.

Illinois State Board of Education. (2023). *Illinois report card data.* Retrieved July 24, 2023, from https://www.isbe.net/ilreportcarddata

Lepkowski, J., Kalton, G., & Kasprzyk, D. (1989). Weighting adjustments for partial nonresponse in the 1984 SIPP panel. In *Proceedings of the Section on Survey Research Methods* (pp. 296–301). Washington, DC: American Statistical Association.

National Center for Education Statistics. (2023a). *Common core of data: Public elementary/secondary school universe survey, 2021–22, provisional version 1a*. https://nces.ed.gov/ccd/files.asp

National Center for Education Statistics. (2023b). *Common core of data: Local education agency universe survey, 2021–22, provisional version 1a*. https://nces.ed.gov/ccd/files.asp

U.S. Census Bureau. (n.d.). *Explore census data*. U.S. Department of Commerce. https://data.census.gov/

Appendix G. Interview Protocol—Original PEAC Members

Question	RQ1: Components	RQ2: Implementation	RQ3: Benefits	RQ4: Challenges	RQ5: Supports
1. Please describe your **role and involvement** as an original PEAC (Performance Evaluation Advisory Council) member. (2 minutes)					
2. What were the **intended benefits** of implementing PERA (Performance Evaluation Reform Act? (5–6 minutes) *Probes:* a. For teachers? b. For administrators? c. For students? d. For school districts?	●				
3. Did PERA have any **goals** other than the intended benefits you just described? (5 minutes)	●				
4. Now that PERA is being implemented, what do you see as the **actual or realized benefits** for educators, students, and others? (5–6 minutes)			●		
5. What has been the **impact** of the legislation on educators and students on . . . (5–6 minutes) a. educator **effectiveness?** b. educator **retention?** c. educator **promotion?** d. student **performance?**			● ● ● ●	● ● ● ●	

Question	RQ1: Components	RQ2: Implementation	RQ3: Benefits	RQ4: Challenges	RQ5: Supports
6. What **challenges** occurred with implementing PERA? (56 minutes) *Probes:* a. Were there any **unintended** consequences or unanticipated challenges? b. What challenges did PEAC members **anticipate**?				●	
7. Please describe the early phases of the **implementation** process of PERA. (5–6 minutes) *Probes:* a. What **issues** came up? b. How were they **resolved**?		●			
8. Do you believe that most districts' systems for evaluating teachers under PERA provide **fair** measures of teacher effectiveness? (56 minutes) *Probes:* a. **What** makes you think so? b. Can you provide an **example**?		●			
9. If **PERA** were revised, what **changes** would you **recommend** that it include? Why? (5–6 minutes)					●

Note. RQ = research question.

Appendix H. Interview Protocol—Early Contributors to PERA Legislation

Question	RQ1: Components	RQ2: Implementation	RQ3: Benefits	RQ4: Challenges	RQ5: Supports
1. What were the intended benefits of implementing the Performance Evaluation Reform Act (PERA) of 2010? (5 minutes) *Probes:* a. For teachers? b. For administrators? c. For students? d. For school districts?			●		
2. In addition to these benefits, what were some of your other goals when proposing PERA and/or Senate Bill 7? (5 minutes) *Probes:* a. Which individuals, organizations, or events played important roles in developing the legislation?	●				
3. What were the actual benefits of implementing PERA, as best as you can describe them? (5 minutes)			●		
4. What is the impact of the legislation on . . . (5 minutes) a. educator effectiveness? b. educator retention? c. educator promotion? d. student performance?			●	●	
5. What challenges did you ***anticipate*** districts might face when implementing the legislation? (5 minutes) *Probes:*				●	

Question	RQ1: Components	RQ2: Implementation	RQ3: Benefits	RQ4: Challenges	RQ5: Supports
a. Were these challenges realized?					
6. Has the legislation had any **unintended consequences** or unanticipated challenges associated with it? (5 minutes) *Probes:* a. Some **examples** might include consequences related to policy, resources, staffing, equity, and local communities. b. Can you briefly **elaborate** on your response?				●	
7. Please describe the **early phases of the implementation** process of PERA. (5 minutes) *Probes:* a. What **issues** came up? b. How were they resolved?		●			
8. What were the **responses to PERA**, particularly from **districts**? (5 minutes) *Probes:* a. How and to what extent did districts **support** the implementation process?		●			●
9. Do you believe that most districts' systems for evaluating teachers provide **valid, reliable measures** of teacher effectiveness? (5 minutes) *Probes:* a. **What** makes you think so? b. Can you provide an **example**?		●			
10. If PERA were revised, **what changes would you recommend** that it include? Why? (5 minutes)					●

Note. RQ = research question.

Appendix I. Focus Group Protocol—PERA Joint Committee Members

Question	RQ1: Components	RQ2: Implementation	RQ3: Benefits	RQ4: Challenges	RQ5: Supports
1. **How long** have you been on your district's PERA (Performance Evaluation Reform Act) joint committee? (5 minutes) *Probes:* a. **When** and **how often** does the joint committee meet? b. **Do most members attend** all meetings?					
2. Based on your understanding, what were the joint committee's **original goals** regarding the evaluation process for administrators and/or teachers? (5 minutes)	●	●			
3. PERA joint committee guidelines state that the group should have equal representation of teachers and administrators. Can you describe **how joint committee membership is determined** in your district? (5 minutes) *Probes:* a. Is there an **application** process? b. Are there **requirements** to be eligible to serve? c. Can anyone who is eligible join, or is there a **limit** to the number of members?	●				
4. What does the **work** of your district's joint committee **involve**? (5 minutes) *Probes:* a. What are the **roles and responsibilities** of joint committee members? What is your specific role? b. What **goals or expectations** have been established for the joint committee? c. How has the role of the joint committee **changed** across time?		●			

Question	RQ1: Components	RQ2: Implementation	RQ3: Benefits	RQ4: Challenges	RQ5: Supports
5. What do you believe are the **strengths** of your district's current system for evaluating educator effectiveness? (5 minutes)			●		
6. What do you believe are the **weaknesses** of your district's current system for evaluating educator effectiveness? (5 minutes)				●	
7. In what ways have you and other members of your district's joint committee been involved in **providing feedback on or making recommendations** to your district's evaluation system? (5 minutes)					●
8. Describe the committee's **process for reviewing** the district's teacher evaluation system, including how often these reviews occur. (5 minutes) *Probes:* a. What **criteria or evidence** does the committee consider when evaluating the district's teacher evaluation system?					●
9. In what ways has your involvement with the joint committee led to **changes** in your district's performance evaluation system? (5 minutes) *Probes:* a. What were the **changes**? b. How were these changes **received** by educators? Evaluators?		●	●	●	●
10. Did the **COVID-19** pandemic play a role in influencing or informing any of these changes? (5 minutes)				●	
11. Please describe any **actions** your district has taken to make the evaluation process more **efficient** and/or less time intensive for educators and evaluators. (5 minutes)					●

Note. RQ = research question.

Appendix J. Focus Group Protocol—Teachers

Question	RQ1: Components	RQ2: Implementation	RQ3: Benefits	RQ4: Challenges	RQ5: Supports
1. Please tell me about your **overall experience** with the district's evaluation process during the 2022–23 school year. (required—9 minutes) *Probes:* a. How was your **impact on student learning** (e.g., student growth) measured or ascertained? (required) b. How was the quality of your professional practice measured? (required) c. What kind of **feedback** did you receive based on observations or student growth? (required)	●	●			
2. Can you provide specific examples of how the **evaluation process** in your district **improved** your professional practice during the 2022–23 school year? (required—9 minutes) *Probes:* a. How did the evaluation process help **identify** strengths or areas for growth? (required) b. What steps or actions did you take to **improve** your professional practice as a result of the evaluation process? (required) c. Are there additional **benefits** of the performance evaluation system for teachers in your district? (required)	●	●	●		

Question	RQ1: Components	RQ2: Implementation	RQ3: Benefits	RQ4: Challenges	RQ5: Supports
3. Can you provide a specific example of how the **district supported** your efforts to improve your professional practice based on the feedback you received as part of your performance evaluation during the 2022–23 school year? (required—9 minutes) *Probes:* a. **Coaching**? (required) b. **Professional development**? (required) c. Time or funding for **peer mentoring**? (required) d. Other **feedback or resources** your district offered you as you sought to improve your instruction? (required)		●	●		
4. To what extent do you feel that your **evaluator** does an **adequate** job in conducting your evaluation? Please elaborate. (required—9 minutes) *Probes:* a. What is the **background** and **experience** level of your evaluator? (required) b. How sufficiently was your evaluator **trained**? (required) c. How **knowledgeable** was your evaluator about what is going on in your professional practice? (required) d. How useful was your evaluator's **feedback**? (required) e. What **additional supports or training** would you recommend that evaluators receive as a part of their preparation to conduct evaluations? Why? (required)		●			●

Question	RQ1: Components	RQ2: Implementation	RQ3: Benefits	RQ4: Challenges	RQ5: Supports
5. How **fairly and accurately** did your district's evaluation process assess your effectiveness as an educator during the 2022–23 school year? (required—9 minutes) *Probes:* a. **Student growth** measures? (required) b. **Observations**? (required) c. **Other measures** or professional practice or impact? (required) d. Why? (as needed, based on participants' responses) e. In what ways is the evaluation process in your district susceptible to **bias or subjectivity**? (required) f. Do you have suggestions for how to mitigate bias or subjectivity (e.g., evaluator training, hiring practice, changes in the measure)? (as needed)	●				
6. What were the most significant **challenges** you faced in relation to your district's performance evaluation system? (required—9 minutes) *Probes:* a. What challenges did you face in terms of being able to **use the feedback** from your performance evaluation to **improve** your professional practice? (required) b. Did you have **sufficient time** to thoughtfully prepare for and engage in the evaluation process? Please explain. (required) c. Were there aspects of the evaluation process that seemed **unhelpful** to your instructional practice? If yes, please elaborate. (required) d. Can you provide a specific **example**? (as needed, based on participants' responses)		●		●	

Question	RQ1: Components	RQ2: Implementation	RQ3: Benefits	RQ4: Challenges	RQ5: Supports
7. How did the **COVID-19 pandemic** impact your performance evaluation, including the process and outcome? (optional—9 minutes) *Probes:* (as needed, based on participants' responses) a. Did the pandemic impact how often you were **observed**? If so, please elaborate. b. Did the pandemic impact the **feedback** you received? If so, please elaborate. c. Did the pandemic impact your **ability to act on the feedback** you received through the evaluation process? If so, in what ways? d. Did the pandemic impact how **student growth measures** were used to assess your effectiveness?				●	
8. What **improvements** to the performance evaluation process would you recommend to your district leaders? To your state leaders? (optional—9 minutes)					●
9. Is there **anything else** that you haven't shared during this focus group that you think is important for us to know for the purposes of this study? (optional—9 minutes)					

Note. RQ = research question.

Appendix K. Focus Group Protocol—School Administrators

Question	RQ1: Components	RQ2: Implementation	RQ3: Benefits	RQ4: Challenges	RQ5: Supports
1. Please tell me about your **role** within the district. Are you a school administrator, an evaluator, or both? (required—2 minutes) (Zoom Poll)					
2. Please describe your **experiences** with your district's teacher evaluation system. (required—10 minutes) *Probes:* a. What **observation tools** do evaluators use to assess aspects of teaching? (required) b. Are evaluation components **weighted** in a way that properly portrays a teacher's overall performance? (required) c. What frameworks or rubrics do evaluators use when observing **teachers**? (required) d. What **software** does your district use to manage the evaluation process? How helpful is it? Are there any barriers to using it? (optional)	●	●	●	●	
3. What type of **training** does your district provide for **evaluators**? (required—9 minutes) *Probes:* a. What are the components of the training? (required) b. Is there an **antibias** component to the training? If so, please describe. (required) c. About **how much time** does it take? (required) d. **How often** is training provided (e.g., one time at the beginning of the year)? (required)		●			●

Question	RQ1: Components	RQ2: Implementation	RQ3: Benefits	RQ4: Challenges	RQ5: Supports
e. How **adequately** did the training prepare you to be an evaluator? (required) f. If not, what **additional training** would be helpful? (optional)					
4. How has the performance evaluation process in your district **supported** improvements to teaching? (required—9 minutes) *Probes:* a. What are the main **benefits** of the performance evaluation system for teachers in your district? b. Can you provide a specific **example of strengths and areas of growth** that were identified for your teachers? c. What **supports** does your district offer to teachers who are seeking to improve? d. What **steps or actions** have you taken to improve the effectiveness of your teachers because of the evaluation process?	●	●	●		
5. How **fairly and accurately** does your district's evaluation process measure teacher effectiveness? Please elaborate. (required—9 minutes) *Probes:* a. Based on **student growth** measures? (required) b. Based on **observations**? (required) c. **Other measures** of professional practice? (required) d. **Why** do you think that student growth, observations, or other measures are accurate or inaccurate measures of teacher effectiveness? (as needed, based on participants' responses) e. In what ways is the evaluation process in your district susceptible to **bias or subjectivity**? (required)		●			

Question	RQ1: Components	RQ2: Implementation	RQ3: Benefits	RQ4: Challenges	RQ5: Supports
f. Do you have suggestions for how to mitigate bias or subjectivity (e.g., evaluator training, hiring practice, changes in the measure)? (as needed)					
6. What are the most significant **challenges** facing leaders and evaluators related to the district's teacher performance evaluations? (required—9 minutes) *Probes:* a. What challenges do evaluators face in **providing useful feedback** to those you evaluate? (required) b. What challenges do teachers face in terms of **obtaining useful** feedback? (required) c. What challenges do teachers face in terms of **improving** their professional practice? (required) d. Did you have sufficient **time** to thoughtfully prepare for and engage in the evaluation process? Please explain. (required) e. Can you provide a **specific example**? (as needed, based on participants' responses)				●	
7. How do you **use information** collected through the performance evaluation system **to improve instructional quality** in your school? (required—9 minutes) *Probes:* a. **Setting goals** for your school? (required) b. Determining **areas** of focus for **professional development**? (required) c. Identifying **supports** for teachers? (required) d. Any other **examples**? (required) e. Do you receive information from the evaluation process in a manner that is **timely** and/or **useful**? (required)	●		●		

Question	RQ1: Components	RQ2: Implementation	RQ3: Benefits	RQ4: Challenges	RQ5: Supports
f. Please describe any **additional** systems of **observation and feedback** outside the performance evaluation system that you use to support the professional growth and development of educators in your building. (optional)					
8. How does your district use the information collected through the performance evaluation system to inform **personnel management decisions** (i.e., tenure or dismissals)? (optional—9 minutes)		●	●		
9. How did the **COVID-19 pandemic** impact performance evaluations in your school? (optional—9 minutes) *Probes:* a. How did the pandemic impact the work of evaluators? (optional) b. Did the pandemic impact the **frequency of teacher observations**? If so, please elaborate. (optional) c. Did the pandemic impact the **quality of feedback** teachers received? If so, please elaborate. (optional) d. Did the pandemic impact teachers' **capacity to act** on the feedback they received through the evaluation process? If so, in what ways? (optional) e. How did the pandemic impact how **student growth** measures were used to assess educator effectiveness? (optional)				●	
10. What **improvements** to the performance **evaluation** process would you recommend to your **district leaders**? To your **state leaders**? (optional—5 minutes)					●
11. Is there **anything else** that you haven't shared during this focus group that you think is important for us to know for the purposes of this study? (optional—3 minutes)					

Note. RQ = research question.

Appendix L. Outreach and Informed Consent

To recruit for interviews and focus groups, the American Institutes for Research® (AIR®) contacted all eligible participants via email. AIR leveraged partners listed in its outreach plan to support these recruitment efforts, and our team provided communication templates for outreach partners to use.

For each qualitative data collection activity, AIR engaged in multiple rounds of outreach. It worked with the Illinois State Board of Education (ISBE) and the Performance Evaluation Advisory Council (PEAC) to select appropriate populations to study and to devise and adjust strategies throughout the outreach period. As part of outreach efforts, informed consent documents were provided as an attachment to the outreach emails, explaining the purpose of the study, the risks and benefits of participating, and how AIR would protect participants' privacy and maintain confidentiality of their data. Prior to conducting interviews and focus groups, AIR also asked participants for verbal consent and permission to record the conversation.

Appendix M. Qualitative Data Analysis Process

A team of four qualitative analysts systematically developed and applied deductive (i.e., pre-established, based on the research questions) and inductive (i.e., emergent, based on the data) codes to all interview and focus group transcripts. AIR created a system of codes to systematically organize and categorize our data, then used NVivo qualitative coding software to apply those codes to key excerpts from interview and focus group transcripts. The analytic process is described below.

Developing deductive codes

Prior to data analysis, two lead qualitative analysts developed a codebook that included parent and child codes. Parent codes refer to high-level codes based on the study's research questions. Child codes are subcategories within parent codes, based on questions from interview and focus group protocols. For example, one parent code based on Research Question 1 is "Components of Performance Evaluation Reform Act (PERA)," and a child code within this parent code is "Student growth measures," as growth measures are one of several evaluation components that participants might discuss. This deductive codebook was developed to establish intercoder reliability on a subset of interview transcripts.

The lead qualitative analysts also developed guidelines in the codebook to provide clarity and consistency in the coding process. For example, the team determined to code primarily at the child code level (meaning more specific child codes over more general parent codes should be prioritized, when possible) and avoid double coding (meaning to assign only one code to a segment of text when possible). To ensure brevity and clarity in the analysis, the guidance included directions to code each segment of text at the minimum amount of text needed to meaningfully interpret the segment as a stand-alone excerpt.

Establishing intercoder reliability using deductive codes

The two lead qualitative analysts trained the rest of the coding team on how to apply the deductive codes to establish intercoder reliability. The qualitative team also met regularly to discuss the coding, review guidelines, and adjust the approach when necessary. For example, during the first meeting, instances where a given code may or may not apply to a text excerpt, as well as instances where double coding might be acceptable, were discussed. The team then used the qualitative coding software to conduct a first round of coding, with all analysts coding the same interview transcript.

After coding the initial transcript, the analytic team used the software to calculate Cohen's kappa (κ), an indicator of intercoder reliability between analysts. More specifically, this statistic

assesses the level of agreement between two or more coders beyond what would be expected by chance (Cohen, 1968). The team calculated Cohen's kappa for every pair of coders, meaning that, for the team of four, six calculations were run. The team continued this process with additional interview transcripts until every pair reached a threshold of at least $\kappa = 0.61$, which indicates substantial agreement between coders (Landis & Koch, 1977). In total, the research team coded four transcripts together to achieve the 0.61 threshold across each pair of coders, using our regular meetings to refine and improve our processes continuously.

Applying deductive codes

After establishing intercoder reliability using the deductive codebook, the coding team independently coded the remaining transcripts. To ensure consistency, the team met on a weekly basis and resolved issues, such as the applicability of a given code or the need to establish additional codes, to characterize a given segment of the transcript more accurately.

Developing inductive codes

While applying the deductive codes, AIR identified additional themes emerging from the data. Coders prepared memos describing themes emerging from the transcripts and areas where the existing codes might require modification or more guidance to better capture patterns in the data and ensure consistency in meanings between coders. The team discussed these memos during weekly meetings. Based on the memos and weekly discussions, the lead qualitative analysts developed a set of inductive codes. The coding team then applied these newly developed codes to a subset of interview and focus group transcripts before finalizing the inductive codebook (Appendix N).

Establishing intercoder reliability using both deductive and inductive codes

After finalizing the inductive codebook, the qualitative analysis team established intercoder reliability when applying all codes to a common data source. The team followed the same process outlined above (i.e., establishing intercoder reliability using deductive codes), ensuring that all pairs of coders met the intercoder reliability threshold ($\kappa = 0.61$).

Applying deductive and inductive codes to all transcripts

After finalizing the codebook and establishing intercoder reliability using both the deductive and inductive codes, the coding team revisited all previously coded transcripts and applied the new, inductive codes as warranted. Also, any remaining transcripts not previously coded were independently coded. Through this process, the qualitative team coded all data sources with all applicable codes as defined in the codebook.

Secondary analysis of coded excerpts

After coding all of the transcripts, the analytic team conducted secondary analysis of coded excerpts to identify patterns within and across qualitative data sources. The team also identified notable outliers, which are salient pieces of information that may not have been part of common patterns. The analytic questions driving this phase of analysis were the following:

- What overall patterns are we seeing in the coded excerpts?
- What within-group patterns are we seeing in the code(s) when reviewed by participant type? By urbanicity of district? By district free- or reduced-price lunch (FRPL) status?
- What trends AND differences are we seeing in the coded excerpts across groups?

The qualitative analysis team met weekly during the secondary analysis phase to discuss emerging findings within and across data sources. Through these iterative rounds of analysis and discussion, the qualitative analysis team identified the key findings shared in the final report.

References

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*(4), 213–220. https://doi.org/10.1037/h0026256

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Appendix N. Codebook

Parent Code	Child Code	Grandchild Code	Definition	Example	Notes	RQ Addressed
Goals and purpose of the Performance Evaluation Reform Act (PERA)			Applied to excerpts that address the purpose or goals of PERA (either originally, or over time)	One purpose or goal is to make evaluations more uniform across the state.		RQ2; RQ3
	Original purpose/goals		Relates to original purpose or goal of statewide educator evaluation system; this includes intended benefits	Originally, reform groups saw PERA as a chance to fire ineffective teachers.		RQ2; RQ3
	Current purpose/goals		Relates to current purpose or goal of statewide educator evaluation system	The state wants to make sure that teachers are receiving ongoing feedback in order to improve so that their students can improve too.		RQ2; RQ3
	What purpose or goals should be		Relates to participant's perception of what the purpose or goal of the statewide educator evaluation system should be	It should be focused on improving teaching and learning.	May be double coded with "Recommendations"	RQ2; RQ3; RQ5
	Removing "bad/ineffective" teachers		Relates to participant's perception that the purpose or goal of evaluation is to get rid of teachers who are bad, ineffective, and so forth	The purpose of PERA was to get rid of the "bad apples."	May be double coded with any other purpose codes	RQ2; RQ3; RQ5

Parent Code	Child Code	Grandchild Code	Definition	Example	Notes	RQ Addressed
			Includes references to the evaluation system being a "gotcha" or a way to "catch" teachers who are not performing well			
	Improving professional practice		Relates to participant's perception that the purpose or goal of evaluation is to support classroom teaching, professional growth, and/or how the educator does their job	The purpose of PERA is to ensure that all districts establish systems and structures so that their educators are able to grow as professionals and be the best they can be for their students.	May be double coded with any other purpose codes	RQ2; RQ3; RQ5
Components of PERA			Applied to excerpts that address a component of the evaluation system			RQ1; RQ2
	Student growth measures		Applied to excerpts that address student growth measures (e.g., state assessments, student learning objectives [SLOs])	Student growth accounts for 30% of the evaluation.	May be double coded as a benefit or challenge	RQ1; RQ2
		SLOs (teacher created)	Applied to specific references to SLOs (i.e., Type III assessments), which are teacher-created assessments used to measure growth in student learning	We use SLOs as a major part of our student growth measures. The welding teacher has different ways to show growth than a math teacher.	May be double coded as a benefit or challenge	RQ1; RQ2

Parent Code	Child Code	Grandchild Code	Definition	Example	Notes	RQ Addressed
		Standardized assessment (not teacher created)	Applied to specific references to district, state, or national standardized assessments used to assess student growth (i.e., Type I or Type II assessments)	If you're a math or English language arts (ELA) teacher, the (NWEA) MAP (Measure of Academic Progress) assessment is used. Everyone else creates their own assessment. Our students' growth is measured using the Illinois Assessment of Readiness (IAR).	May be double coded as a benefit or challenge	RQ1; RQ2
		All-in approach	Applied to excerpts that explicitly state that the district uses an "all-in" approach to assessing the student growth component (meaning everyone uses the same measures, such as MAP growth, even if they are not an ELA or a math teacher, to assess the student growth component)	A few years ago, we adopted an all-in approach. So, everyone's student growth score is measured using the same MAP assessments.	May be double coded as a benefit or challenge	RQ1; RQ2
	Observations of professional practice		Applied to excerpts that address observations of teachers in their classrooms (including references to the observation tool, such as Danielson or Marzano)	My assistant principal usually does the informals, but in terms of formal observations, our department chair does those.		RQ1; RQ2

Parent Code	Child Code	Grandchild Code	Definition	Example	Notes	RQ Addressed
		Charlotte Danielson	Applied to excerpts that specifically reference using Charlotte Danielson's Framework for Teaching (i.e., "Danielson")	We're a Danielson district, but Danielson was never intended to be used as an evaluation tool.		RQ1; RQ2
		Marzano	Applied to excerpts that specifically reference using Marzano's observation framework	We use Marzano.		RQ1; RQ2
	Pre-observation conference		Applied to excerpts that address meetings between teacher and evaluator prior to a formal observation	My evaluator meets with me for about 30 minutes before a formal observation to discuss my lesson and what kind of feedback I'd like to receive.		RQ1; RQ2
	Post-observation conference		Applied to excerpts that address meetings between teacher and evaluator after a formal observation	After my formals, we sit down and discuss the Danielson domains and areas of strength and areas to grow.		RQ1; RQ2
	Other evaluation components		Applied to excerpts that address components of evaluation that are not student growth measures or observations of professional practice; other examples might be artifacts and/or portfolios	In our district, we also use student surveys in our teacher evaluations. That gets calculated into the final score we get.		RQ1; RQ2
	Feedback		Applied to excerpts that address feedback that an educator receives as a	Generally, the feedback I receive from my evaluator is specific and useful.	May be double coded as a benefit or challenge	RQ3; RQ4

Parent Code	Child Code	Grandchild Code	Definition	Example	Notes	RQ Addressed
			result of the evaluation process		May also be double coded with affective codes	
	Professional development		Applied to excerpts that address professional learning opportunities (including coaching) connected to an evaluation	My evaluator noticed that I was really struggling with classroom management, so I started working with a coach in our building and it really helped me to improve.	May be double coded as a benefit or challenge	RQ3; RQ4
PERA joint committee work				Our joint committee meets twice a year and is responsible for . . .		
	Joint committee selection		Applied to excerpts that describe a district's process for selecting joint committee members	We applied to be on the joint committee. Then our superintendent selected the members.		
	Joint committee responsibilities		Applied to excerpts that describe the roles and responsibilities of PERA joint committee members (including expectations to attend meetings)	Our joint committee is expected to attend quarterly meetings and provide recommendations on the evaluation plan.		
	Joint committee processes		Applied to excerpts that describe the joint committee's process for reviewing or improving the district evaluation plan	We review parts of the plan at different meetings. We use a rubric to determine if the plan meets the state's expectations as well as the goals we set out to achieve.		

Parent Code	Child Code	Grandchild Code	Definition	Example	Notes	RQ Addressed
	Updates to district eval plan (past and ongoing)		Applied to excerpts that describe examples of recommendations or changes made to a district evaluation plan by the joint committee	We recommended that the student growth measure component should not only be based on state assessments.		
PERA training			Applied to excerpts that describe training provided to evaluators, school administrators, or teachers related to educator evaluation	The state provides training for all evaluators, and it includes . . .		
	Sufficiency or "dosage" of training		Speaks to whether the training has been sufficient in terms of time or amount	We received training, but it came a little late in the year, and we didn't get a chance to ask all of our questions.		
	Relevance/ content of training		Speaks to whether the training is relevant to evaluators' needs and the requirements of the evaluation system	The training was well aligned to the Danielson Framework and the requirements of our evaluation system.		
	Calibration		Applied to excerpts that address efforts to achieve interrater reliability between evaluators or consistency across evaluators; can be applied to excerpts that describe the need for such training	We watched a video together and all scored it to see if we agreed or disagreed. We saw that we were pretty aligned in how we score teachers using the observation framework.		

Parent Code	Child Code	Grandchild Code	Definition	Example	Notes	RQ Addressed
Impact of PERA			Applied to excerpts that address benefits, challenges, and/or other impacts of PERA	PERA has had several impacts, including		RQ3; RQ4
	For teachers		Relates to PERA's impact on teachers	My teaching has improved since PERA because I've been getting better feedback from my evaluator.	May be double coded as a benefit or challenge	RQ3; RQ4
	For administrators		Relates to PERA's impact on school leaders	PERA has helped to make the evaluation process more uniform, but it has also increased the amount of time I spend filling out paperwork for the people I evaluate.	May be double coded as a benefit or challenge	RQ3; RQ4
	For students		Relates to PERA's impact on students	In cases where the eval process is improving instruction, I guess it's also improving student learning.	May be double coded as a benefit or challenge	RQ3; RQ4
	For districts		Relates to PERA's impact on districts	Teachers and administrators across our district know what's expected of them when it comes to how they'll be evaluated.	May be double coded as a benefit or challenge	RQ3; RQ4
	Time		Relates to the time it takes to implement the legislation as intended and may include things like (1) scheduling, preparing for, and/or completing the	The time and energy it takes to schedule and complete our evaluations by March is just incredible. As a result, it becomes all about compliance	May be double coded with other impact codes	RQ3; RQ4

Parent Code	Child Code	Grandchild Code	Definition	Example	Notes	RQ Addressed
			evaluation process and (2) references to the evaluation process being time consuming or burdensome for teachers or administrators	and checking boxes because we just need to get it finished.	May be double coded as a benefit or challenge	
	Other impacts		Applied to impacts of PERA that do not fit into any other impact subcodes		May be double coded as a benefit or challenge	RQ3; RQ4
Strength/benefit			Relates to a positive impact or strength of educator evaluation	A real benefit of the evaluation system in our district is that everyone knows what's expected of them.		RQ3
	Intended benefit or strength		Relates to a planned or an anticipated benefit or strength	PERA intended to make the evaluation system less subjective and more uniform. That was an intended benefit.	May be double coded elsewhere (e.g., for teachers, for students)	RQ3
	Unintended benefit or strength		Relates to an unexpected or unanticipated benefit or strength	An unintended benefit of PERA is that administrators and teachers came together to agree upon how to conduct evaluations in our district. I don't think that was an intended benefit, but it was definitely an unintended one.	May be double coded elsewhere (e.g., for teachers, for students)	RQ3
	Bias counterexample:		An example or instance of fairness in the evaluation system	Because all of our evaluators use Danielson to observe us, I think it's a fair system in		RQ3

Parent Code	Child Code	Grandchild Code	Definition	Example	Notes	RQ Addressed
	Fairness in system			terms of how I'm observed and the feedback I receive after those observations. I think observations were way more biased when administrators didn't use the same observation tool.		
Challenge			Relates to a negative impact of educator evaluation	A challenge is the student growth component, especially when they use state assessments. There's so much out of a teacher's control that impacts students' performance on those assessments. We shouldn't judge teachers based on those tests.		RQ4
	Implementation challenge		Challenges related to the way the evaluation system is implemented or put into place; can include fundamental belief in evaluating teachers, possibly including using evals for employment decisions	It takes so much of my time to evaluate every teacher in my building, and I'm not sure the time is worth it. We need a way to make the evaluation process more efficient.	May be double coded with other challenge codes (COVID-19 and/or bias) as well as impact codes (i.e., time)	RQ4
	Evaluation component challenge		Relates to whether the evaluation system itself has components or content that causes issues	Creating SLOs is a real challenge. To be honest, I'm not sure the student growth	May be double coded with other challenge codes	RQ4

Parent Code	Child Code	Grandchild Code	Definition	Example	Notes	RQ Addressed
				component is a fair measure of a teacher's effectiveness.	(COVID-19 and/or bias)	
	COVID-19 challenges		Applied to excerpts that address COVID-19's impact on educator evaluation	The pandemic really threw us for a loop because none of our systems were built for a hybrid approach. For example, . . .	May be double coded with other challenge codes (implementation or component)	RQ4
	Bias/unfairness		Applied to excerpts that explicitly address participants' perceptions of bias or unfairness in the evaluation system	Some evaluators use this as an excuse to get rid of teachers they don't like or who are too expensive. It's not based on the educator's effectiveness. It's personal.	May be double coded with impact or challenge codes (implementation or component)	RQ4
Recommendations			Applied to excerpts that address recommendations that could improve the evaluation policy, processes, or supports	I'd recommend that evaluators have no more than 20 people on their caseload.		RQ3; RQ4; RQ5
	Recommended supports or resources		Generally, about supports that districts and schools can provide but may also address state resources; may include references to ways to make the process more efficient or less time consuming	We need more training for evaluators and more support so that they don't have to evaluate so many teachers every year. I recommend using EvaluWise . . . it has been a game changer for our district.	May be double coded with challenge codes	RQ4; RQ5
	Recommended changes to district plan		Changes to district evaluation systems	I'd change the weighting so that student growth was worth only 20% and	May be double coded with	RQ5

Parent Code	Child Code	Grandchild Code	Definition	Example	Notes	RQ Addressed
				professional practice was worth 80%.	challenge or component codes	
	Recommended changes to policy/legislation		Changes to legislation or other policy-oriented changes	I'd take out the student growth measure and just focus on teacher feedback through regular observations.	May be double coded with challenge or component codes	RQ5
	Recommended responses to COVID-19		Applied to excerpts that suggest additional responses to COVID-19's impact on educator evaluation, including things districts did that worked and things participants felt districts should have done	During the fall after the pandemic started, our evaluators did virtual observations. I think that was helpful for us because we learned that you could observe teachers in different ways and still offer meaningful feedback.	May be double coded with challenge codes	RQ5
	Addressing bias/unfairness		Any recommendations specifically to improve issues related to bias or unfairness	To ensure the system is less biased, it's critical that the people designing and monitoring it represent diverse backgrounds and experiences.	May be double coded with challenges codes	RQ4; RQ5
Affective dimensions of evaluation			Applied to excerpts that address the relational aspects of evaluation; the feelings and perceptions associated with being evaluated or with being the evaluator	Relationships are key. If the district admin has a good relationship with the union, if teachers have good relationships with their evaluators, and so on. It makes this so much better than if the relationship is bad or nonexistent.	May be double coded	Varies, depending on excerpt

Parent Code	Child Code	Grandchild Code	Definition	Example	Notes	RQ Addressed
	Trust		Applied to excerpts that address trust or a lack of trust between people	They don't trust us to do our jobs, so they created this evaluation system.	May be double coded	Varies, depending on excerpt
	Consistency		Applied to excerpts that address consistency or lack of consistency in the evaluation process; may refer to consistency across evaluators, consistency in how supports are offered, consistency in how feedback is provided, and so forth	There is zero consistency in the feedback I receive across evaluators. One says that that I'm excellent, and the other says I need improvement.	May be double coded	Varies, depending on excerpt
	Compliance/ checking a box		Applied to excerpts that explicitly state that evaluation (or an aspect of the evaluation process) feels like it is only about compliance or checking a box	This process does absolutely nothing to make me a better teacher. It's all about compliance. Like "check" I did that, "check" I did that. . .	May be double coded	Varies, depending on excerpt
	High stakes		Applied to excerpts that state that evaluation (or an aspect of the evaluation process) feels high stakes	The stakes are really high. I mean, you only get one formal evaluation each year to prove that you're an excellent teacher.	May be double coded	Varies, depending on excerpt
	Anxiety/worry		Applied to excerpts that state that evaluation (or an aspect of the evaluation process) causes fear,	These new teachers are constantly in a state of worry about whether they're going	May be double coded	Varies, depending on excerpt

Parent Code	Child Code	Grandchild Code	Definition	Example	Notes	RQ Addressed
			worry, or anxiety in those being evaluated	to get fired because of a bad evaluation.		
Great quote for final report			Applied to a quote that is particularly illustrative of a theme, concept, or finding; applied to excerpts that could be directly quoted in the final report		Must be double coded	Varies, depending on excerpt
Miscellaneous/to discuss			Applied to excerpts that cannot be coded elsewhere but that align with the research questions and seem important for further analysis			

Note. *Parent code*, *child code*, and *grandchild code* are terms that describe a set of codes that are related. Examples provided are not direct quotes from interviews or focus groups.

About the American Institutes for Research®

Established in 1946, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit institution that conducts behavioral and social science research and delivers technical assistance both domestically and internationally in the areas of education, health, and the workforce. AIR's work is driven by its mission to generate and use rigorous evidence that contributes to a better, more equitable world. With headquarters in Arlington, Virginia, AIR has offices across the U.S. and abroad. For more information, visit AIR.ORG.

AIR® Headquarters
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | **AIR.ORG**

24690_05/24