



Illinois State Board of Education

July 2016

Guidance Document 16-3

PEAC District Evaluation Data Guidance

This document is intended to provide non-regulatory guidance on the subject matter listed above. For specific questions, please contact the person(s) identified in the document.

Dr. Tony Smith, State Superintendent

Printed by AFL-CIO (AFSCME Local #288 and IFSOE Local #3236) Employees

Teacher Evaluation Data Collection

The Performance Evaluation Reform Act (PERA) requires that school districts report basic information about teacher evaluation to the Illinois State Board of Education (ISBE). Those requirements are as follows:

“By no later than June 30 of each year, the State Board of Education shall identify the manner and timeline for the submission of data and other information relative to performance evaluations that each school district must submit. These data and information shall include, but not be limited to, data regarding the performance evaluation rating given to each tenured and nontenured teacher and data about teacher retention, as well as other information specific to the locally adopted performance evaluation plan that will assist the State Board of Education in determining whether performance evaluation systems are reliable and valid, improve student achievement, and contribute to the development of staff.” Illinois Administrative Code (50.130(a))

This document describes data collection and analysis to be carried out at the district level and identifies information that districts must report to ISBE. This document also provides guidance to help joint committees deliver better feedback to teachers and principals and understand the evaluation system, and it is intended to support continuous improvement at both the individual and system levels.

Collecting and analyzing teacher evaluation data at the district level can provide useful information on teachers’ specific needs and areas for improvement in the evaluation system. For example, what if all teachers are receiving high overall practice ratings but low student growth ratings? What if male teachers are receiving systematically different practice ratings than female teachers? What if all nontenured teachers are struggling with the same indicators or domains in their practice? What if all teachers with less than 5 years of experience are showing minimal student growth but teachers with 20 or more years of experience are showing above-average student growth? All of these situations could indicate a need for improvement in measures, point to particular next steps in terms of professional development or supports needed for teachers, or simply indicate a need for additional analysis.

This guidance document is intended to accomplish the following:

1. Help joint committees consider what kind of data to collect, how to collect it, and what to do with it; and
2. Help joint committees do some proactive planning about some potentially common scenarios in data and how to address them.

Fixsen et al. (2005) have identified six stages involved in implementing an initiative, e.g., a teacher evaluation system. Figure 1 shows when a district implementing teacher evaluation may reach each of these stages and what data collection and analysis tasks are appropriate at each stage.

Figure 1. Possible Stages of Implementation Timing and Data Collection and Analysis

| Year | Implementation Stage | Data Collection and Analysis |
|-----------|---|---|
| Planning | Exploration and Adoption: Initial phase of implementation—identify the need for an intervention, gather stakeholder support, and choose an intervention. | Plan for data collection. |
| Years 1–2 | Program Installation: Prepare for installation without changing practices, including gathering resources and training. | Focus on data needs and collection. Analyze data to: <ul style="list-style-type: none"> ▪ Monitor the fidelity of implementation; ▪ Examine educator perceptions; and ▪ Monitor the alignment of ratings. |
| | Initial Implementation: Start to implement a new strategy, policy, or program and confront fears and uncertainty that occur with any change. | |
| Years 3–5 | Full Operation: The new program is fully integrated into practice, gradually becoming accepted practice. | Analyze data to: <ul style="list-style-type: none"> ▪ Monitor the distribution of ratings and ▪ Measure system impact. |
| | Innovation: Opportunities for refining the practice and additional customization occur during the innovation phase. | |
| | Sustainability: Ensure that the implemented practices are continued through staffing and funding changes and sustained community and political support. | |

Districts typically engage in program installation and initial implementation in the first and second years of implementation of an evaluation system. During that time frame, a joint committee may not have the time or resources to deeply analyze evaluation data because other aspects of implementation take priority. Thoughtful collection of the data at the beginning of implementation, however, can allow for analysis when the district moves into later stages of implementation, such as full operation, innovation, and sustainability.

This guidance document describes five purposes of data collection, with guiding questions and data sources for each purpose that can help inform joint committees. It is important to note that answers to the guiding questions may necessitate further investigation into the root causes of those answers to determine whether or not the system is being implemented as intended or is achieving desired outcomes. Data interpretation needs to take place in the context of the setting. Throughout this document, possible scenarios that a district may encounter while analyzing evaluation data are described. The guidance document concludes with a discussion of data elements to collect as well as other considerations.

It is important to note that answers to the questions that follow are not absolute but, rather, suggest possible approaches. Responses to data collection questions should lead to further investigation into whether or not the system is working as intended.

Purposes of Data Collection and Analysis

Prior to collecting any data, joint committees should determine the purposes and desired use of data collection and analysis outcomes.

Purposes. Specifically, joint committees should consider which individuals or groups will use the collected information to make decisions and what types of decisions they will make. Joint committees should not collect data that are not going to be analyzed. The use of outcomes should drive the data collection.

Timing. Joint committees should consider what data are going to be collected regularly and what extra data needs to be collected. By considering these issues early and allocating sufficient time and resources, the joint committee can ensure that it has the necessary data for its intended purposes.

Analysis. When analyzing data over time, a joint committee should consider patterns and trends. Decisions or changes to the system should not be made until patterns or trends are discovered.

Short Term: Years 1 and 2

Monitoring Fidelity of Implementation

This type of data collection is intended to inform district leadership and those implementing evaluation systems, including joint committees, about the extent to which the system is being carried out as intended. These types of data can help inform decisions about the need for additional training or communication or for changes to the measures or system itself. Specific questions that can be addressed include the following:

- Are all schools implementing the evaluation system according to the district timeline and within district parameters? For example, were all observations completed within the expected time frames? Were all expected, pre-observation and post-observation conferences held, and did they take place within the anticipated time frames?
- Are measures being accurately implemented? For example, do observational ratings align with the evidence cited for them? Do approved student learning objectives (SLOs) contain all of the information required by the district's SLO system/templates? Have summative ratings been accurately computed based on joint committee decisions?
- Were evaluators certified or recertified in the expected time frame? How many evaluators were certified or recertified on their first attempt, and how many attempted certification more than once?
- Do patterns or relationships exist between evaluators and ratings? For example, do some evaluators consistently assign higher ratings than others?
- What have been the costs of implementing the evaluation system? Do these align with expectations?

Potential data points that a joint committee could collect and analyze to monitor the fidelity of implementation include information about the primary and secondary evaluator for each teacher,

dates when observations and other related processes took place, costs for training or other implementation activities, and/or dates when student growth measures were submitted and approved.¹ Joint committees could also consider sampling observational or student growth data to check for accuracy.

Examining Educator Perceptions

This type of data collection is intended to inform joint committees and those implementing evaluation systems whether educators think the evaluation system is helpful. If educators do not believe in their evaluation system, then it may be limited in its ability to improve teacher practice and, ultimately, student learning. Thus, information about educator perceptions could point to changes that may be needed in communications, in supports, or in the system itself. Specific questions that can be addressed include the following:

- Do teachers and evaluators believe that the training, communication, and/or supports provided to implement the evaluation system are adequate? What are barriers or challenges to implementation and how could these be resolved?
- Do teachers and evaluators believe that the evaluation system is helpful and is making a difference? Which elements of the evaluation system are most or least useful?

Data can be collected through surveys, focus groups, or interviews, depending on the depth of information desired and the capacity to carry out data collection. Joint committees may also want to capture background data about respondents to look for patterns—for example, whether more experienced teachers have a particular perception and less experienced teachers have a different perception.

Monitoring Alignment of Ratings

This type of data collection is intended to inform district leadership and those implementing evaluation systems, including joint committees, about how well various measures align, in order to inform decisions about the need for additional training or research on measures. It is not necessarily true that all measures should align precisely at an individual teacher level—that is why we use multiple measures—but, in the aggregate, we expect that teachers who demonstrate strong performance on one measure (e.g., observations) would also score well on others (e.g., student growth) (Cantrell & Kane, 2013). The extent to which such relationships exist can be affected by the accuracy and reliability of the measures themselves, such that lack of alignment in ratings could indicate an issue worth investigating further. For example, lack of alignment could indicate that observational or student growth data are not accurate. Specific questions that can be addressed include the following:

- What is the correlation between professional practice and student growth ratings?
- Are professional practice and student growth ratings generally similar within schools and across the district?

¹ Potential data points for each data collection purpose were adapted from Chatis, C. & Stanton, L. (2013, October 13). *Quality evaluation rollout data analytics convening* [PowerPoint slides]. Fairfax, VA: Reform Support Network.

- Do teacher growth targets reflect student growth trends, if historic student achievement data are available? For example, if students have historically shown high amounts of growth in a course, do teacher growth targets align with these levels of growth?

Potential analyses that a district could carry out to monitor the alignment of ratings include the percentage of teachers with growth and practice ratings that are the same, or within one level of each other. Districts could also look for cases with significant lack of alignment.

***Scenario 1:** What if a joint committee finds a discrepancy between teacher practice scores and student growth scores? For example, teacher practice scores may be consistently higher or lower than student growth scores. This type of result may indicate a challenge related to any of the following issues:*

- **Evaluator training.** It is possible that further training is needed to improve the evaluator’s use of the observational tool. For example, evaluators may need additional training and support on using the professional practice rubric with teachers of certain subjects. The Performance Evaluation Advisory Council (PEAC) guidance document [Guidance on Building Teacher Evaluation Systems for Teachers of Students With Disabilities, English Learners, and Early Childhood Students](#) describes some best practices for this type of situation.
- **Teacher training.** Teachers may need training or practice in analyzing student data and setting student growth goals. Without this training and practice, teachers may be setting student growth targets too high or too low. Also, teachers may need additional training and support in analyzing student data and setting student growth goals for specific groups of students, taking into account student baseline scores. If teachers are writing their own assessments to measure student growth, they may need training on writing a quality assessment.
- **Instruments.** The joint committee should consider whether or not the professional practice rubric, student growth assessments, and student growth guidelines are capturing the correct information.
- **Bias.** Evaluators may be showing unconscious biases toward certain groups of teachers. Evaluators may benefit from training and support in countering their implicit biases (for example, see Gates Foundation 2013; Whitehurst, Chingos & Lindquist, 2014; Ross, 2016).

***Scenario 2:** What if growth targets on student learning objectives are not being set appropriately?*

If joint committees analyze data and determine that teachers are setting growth targets that are not realistic and attainable and that evaluators are approving those growth targets, they should consider the following:

- **Teacher and evaluator training.** ISBE has several resources that can support districts, evaluators, and teachers in creating and approving SLOs. The ISBE [Student Learning Objective Guidebook](#) provides information on setting growth targets. In addition, online ISBE modules on [growth targets, outcomes, and teacher ratings](#) and [guidance for special populations](#) are available. All ISBE SLO resources are available at <http://www.isbe.net/assessment/htmls/balanced-asmt.htm>.

Long Term: Years 3–5

Monitoring Distribution of Ratings

This type of data collection is intended to inform district leadership and those implementing evaluation systems, including joint committees, about how well measures differentiate among teachers. Lack of differentiation, or patterns of differentiation that vary within a district, could indicate an issue with the accuracy of measures that could necessitate further training, further analysis, or adjustments to measures themselves. Specific questions that can be addressed include the following:

- How many teachers are rated at each level of performance? What components overall are highly rated? What components overall have low ratings?
- How are ratings distributed across schools, grade levels, and subjects?
- How are ratings distributed across teachers with different characteristics, such as background (e.g., gender, race/ethnicity), experience or preparation (e.g., years of experience, type of degree or preparation)? For example, do teachers with more experience receive systematically higher or lower ratings than less experienced teachers?
- How are ratings distributed across teachers with different classroom characteristics? For instance, do teachers who teach many English language learners, students with disabilities, chronically absent students, or other types of students receive systematically higher or lower ratings than teachers with fewer of these types of students?
- How are ratings distributed for teachers using different measures or assessments for student growth? Do the student growth guidelines seem to be working?

Potential data points that a district could analyze to monitor the distribution of ratings include basic information about teachers such as background, grade, and subject taught; experience; and preparation. Combining this information with ratings data could provide useful insight into patterns of performance or potential challenges with measures.

Scenario 3: What if there is large variability in ratings from year to year?

If a joint committee observes variability of scores from year to year, it should consider the following:

- **The sample of teachers being evaluated.** PERA allows for tenured teachers to be evaluated biennially (every other year). The joint committee should look at the ratings of individual teachers over time. Are the scores of teachers fluctuating, or are the scores of teachers consistent with differences between the teachers who are evaluated each year?
- **Evaluator training.** The joint committee should make sure that all evaluators have passed the state certification for conducting evaluations. The district should also consider what kind of refresher training evaluators need annually to refine their skills and how the district can provide that support.

Scenario 4: *What if there is a relationship between contextual or out-of-school factors (e.g., student attendance) and teacher ratings?*

After analyzing data, the joint committee may determine that the current evaluation system does not account for contextual factors in the district. Or, perhaps the context of the district has shifted since the evaluation system was created. If either of these is the case, the district should consider the following:

- **Instruments.** The joint committee should consider whether the professional practice rubric, student growth assessments, and student growth guidelines are capturing the correct information. The [Guidance on Creating Operating Guidelines for Student Growth Models in Teacher Evaluation Systems](#) document offers suggestions that the district may find helpful.
- **Bias.** Evaluators may be showing unconscious biases toward certain groups of teachers. Evaluators may benefit from training and support in countering their implicit biases (for example, see Gates Foundation 2013; Whitehurst, Chingos & Lindquist, 2014; Ross, 2016).
- **Student growth guidelines.** The joint committee should consider whether or not current business guidelines regarding student attendance, student growth, and growth targets need to be revised. The districtwide teacher evaluation data can provide the necessary information for making informed decisions about these and similar topics. Also, [Guidance on Creating Operating Guidelines for Student Growth Models in Teacher Evaluation Systems](#) includes information and prompts that the district may find helpful. Topics that the district might consider include student attendance; for example, should the threshold for including students in growth measures be raised or lowered based on district attendance patterns?

Measuring System Impact

This type of data collection and analysis is intended to inform joint committee members and those implementing evaluation systems about the extent to which the evaluation system is doing what it is supposed to be doing. For example, the evaluation system may be intended to inform decisions about professional development, placement, mentoring, and granting tenure. To know whether or not the system is achieving these goals, districts must gather information on whether data are being used for these purposes. The evaluation system may also be intended to improve teacher practice. To determine whether the system is meeting this goal, districts must gather and analyze information about whether or not teacher practice is changing. Because this purpose requires the evaluation system to be fully implemented, analysis of data on a system's impact cannot be completed during the first few years of implementation. Specific questions that can be addressed include the following:

Individual impact

- How do individual teachers, principals, and district staff use teacher practice or student growth data from the evaluation system, if at all? For example:
 - Do teachers use the information to write professional learning plans or request support?

System impact

- How do individual principals and district staff use teacher practice or student growth data from the evaluation system, if at all? For example:
 - Do principals use the information to identify potential teacher leaders or mentors, to make classroom assignment decisions, to assign coaches or peer supports, or to make recommendations about tenure or other personnel decisions?
 - Do district staff use the information to determine professional development needs or offerings for the district, to evaluate professional development or other programs, or to make tenure or other personnel decisions?
- If teachers, principals, and district staff do not use information from the evaluation system as intended, why not? What are the biggest challenges or obstacles to using data as intended?

Data to address these questions could be collected through surveys, focus groups, or interviews, depending on the depth of information desired and the capacity to carry out data collection. Joint committees could also match data about roles such as mentorship or teacher leadership roles with evaluation data to determine whether or not a relationship exists between ratings and these types of roles, or they could match data about professional development attendance with evaluation data to examine the relationship between teacher practice data and identified areas of strength or weakness.

Data Elements to Collect

Joint committees may need to prioritize how much and what data to collect and analyze based on their time and capacity. Some of the purposes require simply collecting or analyzing data that should already exist as part of system implementation or as part of other district information systems. For example, teacher ratings are produced as part of the evaluation system, and it may simply be a matter of systematically collecting or organizing these data. Similarly, teacher background data (e.g., experience or assignments) may reside in a district's human resources information system. In those cases, districts may merely need to develop and carry out an analysis plan. In doing so, district size may be a consideration; for example, small districts may be limited in their ability to draw definite conclusions based on analysis of data for small numbers of teachers or evaluators. In other cases, new data (such as information on educator perceptions) may need to be collected through surveys, focus groups, or interviews.

At a minimum, PEAC recommends that joint committees prioritize collecting and analyzing data that focus on monitoring the fidelity of implementation and examining educator perceptions. Collecting data for these purposes may also provide useful information for other purposes. Data needed to address the purposes described in this document likely include the following elements:

- **Teacher rating data:** Observation ratings by indicator, overall professional practice rating, individual measure of student growth ratings, overall student growth rating, summative rating data
- **Evaluator data:** Certification or recertification status and dates of certification or recertification

- **Teacher background data:** Years of experience, school, grade, subject, and classroom assignment; tenure status; gender, race/ethnicity; qualifications or credentials; type of professional development attended
- **System implementation data:** Dates of observations (plus pre- and post-conferences, if relevant) and possibly samples of SLOs or student growth calculations
- **Perception or system use data:** Responses to interview, survey, or focus group questions about utility and accuracy of the system and data and uses of data

Other Considerations

In developing data collection, analysis, and reporting plans, districts should consider minimum sample sizes for analysis and reporting, rules or policies regarding confidentiality or personally identifiable information, and laws and/or regulations regarding open records or freedom of information.

Confidentiality and Data Security. When determining how data will be analyzed, joint committees should consider the confidentiality of the data. Will data be personally identifiable at the individual teacher or evaluator level? What instructions and parameters will be given to those collecting or analyzing data to ensure that personally identifiable information is protected and used appropriately?

Analysis and Reporting. Joint committees should also consider what data can be appropriately or legally analyzed or reported. For example, if a district has only a few teachers or evaluators with a particular characteristic (e.g., more than 10 years of experience), then analyzing or reporting data disaggregated by those characteristics could prove misleading or jeopardize confidentiality. Joint committees may also need to consider what data can be made public by request.

References

- Cantrell, S., & Kane, T. J. (2013, January). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study*. Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from http://k12education.gatesfoundation.org/wp-content/uploads/2015/05/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf
- Chatis, C. & Stanton, L. (2013, October 13). *Quality evaluation rollout data analytics convening* [PowerPoint slides]. Fairfax, VA: Reform Support Network. Retrieved from http://marylandpublicschools.org/MSDE/programs/tpe/docs/QER_DataDashboardConveningPresentation.pdf
- Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of Southern Florida. Retrieved from <http://ctndisseminationlibrary.org/PDF/nirnmonograph.pdf>
- Jiang, J. Y. & Sporte, S. E. (2016). *Teacher evaluation in Chicago: Differences in observation and value-added scores by teacher, student, and school characteristics*. Chicago: University of Chicago Consortium on School Research. Retrieved from <https://consortium.uchicago.edu/sites/default/files/publications/Teacher%20Evaluation%20in%20Chicago-Jan2016-Consortium.pdf>
- Ross, H. J. (2016). *Everyday bias: Identifying and navigating unconscious judgments in our daily lives*. Lanham, MA: Rowman and Littlefield.
- Whitehurst, G. J., Chingos, M. M., & Lindquist, K. M. (2014, May). *Evaluating teachers with classroom observations: Lessons learned in four districts*. Washington, DC: Brown Center on Education Policy at Brookings. Retrieved from <http://www.brookings.edu/~media/research/files/reports/2014/05/13-teacher-evaluation/evaluating-teachers-with-classroom-observations.pdf>