

Performance Evaluation Advisory Council (PEAC)

April 19, 2013

Meeting Minutes Illinois State Board of Education Building, Springfield, Illinois

Members Present

Kristen Adams, Benjamin Boer, Angela Chamness, Dawn Conway, Paula Crane, Don Daily, Dr. Randy Davis, Dr. Gail Fahey, Kurt Hilgendorf, Dr. Steven Isoye, Dr. Joseph Matula, Vicki Phillips, Dr. Stephen Ponisciak, Dr. Diane Rutledge, Michelle Standridge, Janet Tate (representing Audrey Soglin), Dr. Linda Tomlinson, Dr. Richard Voltz, Paulette Poncelet (replacing Matthew Lyons).

Observers Present

Amy Alsop and Kathy Shaevel (Illinois Federation of Teachers [IFT]); Christi Chadwick, (Illinois State Board of Education [ISBE], on phone); Dr. Steven Cantrell (Bill & Melinda Gates Foundation); Lisa Hood and Dr. Erika Hunt, (Illinois State University, Center for the Study of Education Policy [CSEP]); Mary Jane Morris and Mary McDonald (Consortium for Educational Change [CEC]); Kellee Sullivan (ISBE); Bradford White, Illinois Education Research Council [IERC].

Facilitators Present

Cassandra Meyer, Larry Stanton, Nick Pinchok, and Meghan Zefran (American Institutes for Research [AIR])

Meeting Called to Order

The meeting was called to order at 10:10 a.m. by Dr. Linda Tomlinson.

I. Welcome, Announcements, and Updates¹

A. *Growth Through Learning* Update

Mary Jane Morris (CEC) provided a written report on completion rates for the teacher and principal evaluation training. Ninety-two percent of teacher evaluators have completed Module 4, and 56 percent have completed Module 5. Ninety-seven percent of teacher evaluators are proficient in Module 2. Three individuals have exhausted all of their attempts to pass this observation test and cannot go any further. About 83 percent of principal evaluators have completed Modules 1–3. Individuals at many institutions of higher education have entered into the system. Most user licenses for the Teachscape portion of the training were set to expire on May 30, 2013; CEC has negotiated with Teachscape to extend those licenses for six months.

A PEAC member asked what would happen during the summer. Morris responded that CEC's contract ends June 30, 2013.

When asked if any teachers had taken the training outside of their districts, Morris responded in the affirmative.

¹ The agenda shifted during the meeting, so events are reported in the order in which they occurred.

In the legislation, evaluators are required to recertify once every five years. Vicki Phillips encouraged members to communicate to the field that the user licenses have been extended so that observers can recalibrate if they want to do so.

B. Race to the Top Update

Ms. Phillips provided an update on Race to the Top. Three requests for sealed proposals (RFSPs) in the fiscal department are waiting to be released. One RFSP is for a learning assessment system so that districts can work together on assessments and items by content-area groups. Another RFSP is related to induction and mentoring. The third RFSP is a renewal of the evaluation training work. In addition, there may be a teacher preparation project related to integrating Common Core State Standards and a Performance Evaluation Reform Act (PERA) evaluation contract with Westat.

C. Communications Update

Dr. Diane Rutledge shared that she testified at the Legislature about funding for PEAC. She added that the reception was more favorable than it was last year.

Dr. Rutledge also stated that PEAC member organizations need to make sure that important information gets out to the field. PEAC members explained how they are or will be sharing information from PEAC with their members.

Dr. Rutledge reminded members to make sure they include PEAC information at their conferences. Some members already have responded about the ways they plan to communicate. Dr. Philips asked for other conference information.

When one member suggested that all conferences use the same presentation, Dr. Philips indicated that there is an updated version of the original PEAC communication PowerPoint presentation and that it needs to be posted on the PEAC website.

Larry Stanton clarified that the main goal is to drive people to the website and guidance. When a member asked if there is a fact sheet that includes the information from the PowerPoint presentation, Dr. Joseph Matula stated that he has one that he uses with his classes. He offered to send the document to ISBE and said that AIR will make sure it looks similar to the other guidance documents.



Guidance

III. Stage 1: Introduce

A. MET Study Presentation (via Webinar)

Dr. Steven Cantrell, Chief Research Officer at the Bill & Melinda Gates Foundation, provided an overview of the Measures of Effective Teaching (MET) study. This webinar focused on highlights from the most recent report because Dr. Cantrell already has presented to this group about earlier reports.

The research found that 85 percent of teachers are indistinguishable from average. Instead of the distribution of teachers being shaped like a bell curve, the curve of teacher effectiveness included most teachers in the center near average, with few outliers at the high and low end.

Dr. Cantrell discussed what the research showed about each one of the following three research questions.

Research Question 1: Can we measure effective teaching with confidence? The answer is yes.

- The MET project used value-added student growth data, student surveys, and observations to determine what part of the results was from teacher effect and what part was from different types of students. Statistically, it is possible to control for issues such as socioeconomic status (SES), race, or gender. The project did a random assignment of classes of students to teachers. This random assignment allowed the study to control for student motivation.
- In addition, the study looked at the following questions: To what extent does effectiveness follow the teacher? When there was a random match, did effective teachers stay effective? Did low-performing teachers stay low-performing? Dr. Cantrell said that the prediction was very close to what was measured. At the extremes, the predictions were conservative (great teachers showed higher student gains than expected, and poor teachers showed lower student gains than expected).

Research Question 2: What weighting of measures provided the most reliable results?

- Four different options were tested.
- Equal weights on value-added, student surveys, and observations provided the most reliable results.

Research Question 3: What mix of peer and administrator observations are the most reliable and also remain feasible?

- The study found that there are many roads to reliability. The “sweet spot” in terms of observers is two administrative observers, each observing the teacher for 45 minutes; or one administrative observer observing for 45 minutes, with three peer observers observing for 15 minutes each.
- Both options showed 67 percent reliability.

Responding to member questions, Dr. Cantrell clarified that the observation reliability used in other parts of the study (.67) were comparable to these two options. When an observer had background knowledge about a teacher, it increased reliability.

The Bill & Melinda Gates Foundation is working to build a field guide, in addition to the framework for improvement-focused teacher evaluation systems that it already produced.

Three recommendations related to teacher quality are being considered. These three recommendations relate to measures, validity, and investment in improvement.

It is important to make meaningful distinctions about investing in improvement. Most of the policy conversation has been around quartiles of teachers; that division is not what was observed in MET. Teachers should be in different categories only if their impact on students is demonstrably different. States and districts should prioritize support and feedback. States and districts should use data for decisions at all levels, including judging the effectiveness of support of teachers.

About 85 percent of teachers were indistinguishable from average; but, for the sake of improvement, it is important to identify why the teacher scored so low. A teacher might need different supports, depending on whether the teacher was low on student surveys, observations, or value-added scores.

Stanton asked: How did having four ratings align with this idea? Dr. Cantrell answered that most teachers would be rated as *basic* or *proficient* on Charlotte Danielson’s *Framework for Teaching*.

Stanton then asked if the teacher spread was the same for test data. Dr. Cantrell responded that test data might exaggerate the spread of teachers.

IV. Lunch

II. Approval of March Minutes

Paulette Ponsel moved to approve the minutes. Benjamin Boer seconded the motion.

Ms. Phillips requested that when people give comments on guidance documents and send them to AIR they also should cc her on the e-mails so that ISBE can track the feedback related to the guidance documents.

Guidance

III. Stage 1: Introduce

B. Evaluation of Early Childhood Teachers

Lisa Hood and Dr. Erika Hunt (CSEP) presented their work on the evaluation of early childhood teachers. Their organization just merged with the National Board Resource Center.

Hood stated that *early childhood* refers to prekindergarten through Grade 3. Although the MET study validated the *Framework for Teaching* for teachers from Grade 4, CSEP just submitted a proposal to validate the instrument at the prekindergarten through Grade 3 level.

CSEP looked at the *Framework for Teaching* and identified new examples to align with prekindergarten through Grade 3 teacher roles. Then CSEP developed and coded videos, using the adapted *Framework for Teaching*. CSEP is developing training for principals and teachers.

CSEP gathered feedback from early childhood stakeholders in April 2013 and will do so again in May 2013. National feedback will be solicited from the National Association for the Education of Young Children conference in June 2013. A pilot on the revised *Framework for Teaching* is scheduled for fall 2013, involving the principal and two teachers in each of seven schools throughout the state.

The professional development for the pilot will include three modules similar to the *Growth Through Learning* modules. Module 1 will focus on why a pilot is needed, understanding the framework, and exploring examples at different grade levels. Module 2 will focus on understanding the professional improvement cycle and practicing coding videos. Module 3 will examine level-specific artifacts of practice across all four domains.

CSEP also just submitted a proposal on the development of student/parent surveys and student growth sources. Hood noted that they were thinking of using the Kindergarten Individual Development Survey (KIDS) as a starting point.

A member asked to what extent Charlotte Danielson agreed to changes to the *Framework for Teaching*. Hood clarified that Danielson approved the process. CSEP cannot change the framework or critical attributes, but the organization can add supplemental examples.

Members asked about whether similar revisions could be made for special education teachers and English language learners. Hood clarified that Danielson would allow the revisions. A member clarified that

external funding would be needed. Hood noted that their biggest expense is the stipends for practitioners and staff time.

A member asked if the *Framework for Teaching* is part of the state default model. Dr. Linda Tomlinson stated that the guidance from PEAC would be to use the *Framework for Teaching*, but the framework is not officially part of the default model. Ms. Phillips stated that PEAC vetted a number of resources for measuring practice. Through its vetting process, the tool most recommended is the Danielson *Framework for Teaching*. If a joint committee cannot come to agreement, the committee should adopt this framework. Members questioned what a joint committee would do if the members could not come to agreement on the practice side of the teacher evaluation system. Dr. Tomlinson stated that she would bring back an answer to this question from the general counsel.

A member asked if there will be new training for kindergarten through Grade 3 principals. Ms. Phillips clarified that the work by CSEP will inform the field, but that work is not ISBE work.

Ms. Phillips expressed concern about using the KIDS assessment in a way that impacts student growth, and she stated that CSEP should take a broad look at that issue. Hood stated that CSEP would ideally convene a steering committee for the work, but the organization has not started those conversations. Hood asked for suggestions for experts to include on the committee. A member recommended that CSEP consider the existing Kindergarten Readiness Stakeholder Committee on reporting, and another member suggested considering members from the task force from the state board piloting the new early learning standards.

SM: Require

V. Stage 3: Decide

A. Evaluating Teachers of Special Education Students (Guidance)

This presentation, which focuses on evaluating teachers of special education students, was postponed until the May 2013 meeting.

VI. Decide

A. State Model for English Language Arts (ELA)/Mathematics for Type I Assessments

Stanton recapped the work that has been done since the last meeting. There was a conversation among Dr. Stephan Ponisciak, Benjamin Boer, and experts from AIR.

Stanton clarified that the purpose of this discussion is to do the following:

1. Develop a shared understanding of adjusted growth, typical growth, and measurement error.
2. Understand two approaches for measuring growth—individual student against a target and growth model.
3. Decide whether PEAC wants to recommend that ISBE assess the feasibility of one or both approaches. ISBE will report back next month.

A member questioned whether a decision had been made about the assessments that are part of the default model. Dr. Ponisciak clarified that the Measures of Academic Progress (MAP) and ACT's Educational Planning and Assessment System (EPAS) were the two assessments that had been agreed upon as part of the state model, although the exact years that those assessments would cover had not been determined.

Michelle Standridge said that the Illinois Education Association does not want to use Type I assessments as part of the state model.

Boer noted that there was some confusion and concern at the last meeting. He reiterated that assessments are part of a multiple-measures system. As Dr. Cantrell noted in his presentation, the error on different measures does not correlate. If a district uses two classroom-based assessments, there are error correlates and overall reliability is lowered.

Boer outlined the principles for the state growth-model system that he and Dr. Ponisciak agreed upon as a starting place.

Following are the principles for the state growth model:

1. Multiple measures
2. Minimum complexity
3. Minimum requirement: adjusted growth model
4. Moving districts to better use assessments
5. Minimum requirements: adjusted for measurement error
6. Continuous improvement approach
7. Huge step from attainment (current culture)

Dr. Ponisciak stated that any growth system must use an adjusted growth model. Adjusted growth models account for student starting points. Expected growth for each student is identified. Boer stated that it cannot be assumed that lower performing students will grow more than higher performing students.

A member asked if adjusted growth is potentially discriminating. Does it, at some level, lead to having lower expectations for low SES students or students of a certain race?

Stanton stated that although SES might have affected prior performance, an adjusted growth model assumes the same ending point. Boer stated that the point is valid, but he is not overly concerned about including low expectations in the future discussion of measurement error and their broad proposal for addressing it. At the end of the day, this proposal is the most fair. A member asked if the adjusted growth model is different from other models. Dr. Ponisciak stated that he would not even consider adjusted growth to be a full model, just a starting principle.

Dr. Ponisciak shared an example of what growth could look like. Boer clarified that two concepts are being illustrated: adjusted growth and measurement error. Measurement error does not permit anyone from saying with certainty that a student made typical growth if the student's score from the ACT Plan assessment moved from 9 to 10 because the measurement error for this test is from 0 to 4. Stanton clarified that although this example is an EPAS one, the same concept applies to MAP.

Two members asked if there was research about a teacher's impact on the growth of a student's score on the EPAS assessments. Boer responded that the question would have to be taken to ACT.

A member restated the previous concern: that the system is setting low expectations for low performers.

Boer stated that by the time a student comes to high school, if the student scores a 12 on the Explore test, it is clear that this student has not been successful by that measure. Because of this starting place, there

are reasons to believe that that student would not be as successful moving forward. By looking at averages and national norms, this result might be seen. However, there are students who move a great deal on this measure.

Boer also stated that the other piece of that question relates to the tests themselves. There is no natural score for learning fractions. Tests have to try to measure that learning. The member agreed that the expectation should be the same for all students who come in with the same starting score.

A member asked how this concept fits with state standards and expectations for knowledge at a grade level. Boer stated that specific expectations were the purpose of the Common Core State Standards; but, in reality, students enter the classroom with different levels of knowledge.

Dr. Ponisciak noted that it is impossible to obtain an exact measure of what a student knows. Tests are essentially polling what a student knows. He explained the concept of reliability and of measurement error.

Concept 1. Boer described the first of two concepts being proposed to PEAC. The first concept has the least technically heavy lift and attempts to be conservative about how to use the growth metric. In the first concept, the result will be the percentage of students who made expected growth for each teacher. Three things can be said about a student's performance: (1) The student made expected growth (surpassed historical typical gain by at least the error), (2) the student did not make expected growth, or (3) it could not be determined if the student made expected growth. Because of measurement error, most students will be in the third category.

After answering some clarifying questions, Boer distributed a Northwest Evaluation Association (NWEA) MAP report. (*Note:* The report is not to be distributed outside of this meeting.) Boer described data in the columns of the NWEA report. NWEA recommends that a state using NWEA MAP to measure teacher impact on student growth should create a state database of students who took the MAP assessment; it does not recommend that the nationally normed expected data be used in teacher evaluation. Boer stated that the expected growth from an Illinois database likely would be similar to a national database.

A member asked about the feasibility of a district creating its own norms. Dr. Ponisciak stated that although that process might work for a big district, it would be problematic for a small district without historical data.

Boer provided an example of what Concept 1 might look like for a teacher.

Members asked clarifying questions, and Boer explained that the typical error of both pretests and posttests is included in the measurement error. He further clarified that in Concept 1, the system would not say that a teacher is unsatisfactory if one student missed growth. Further work of this group would be to decide an acceptable percentage of students who missed growth.

A member asked what percentage of students would be likely to fall into that "not sure" category. Boer answered that without historical data, he did not know. Dr. Ponisciak stated that it might be from 50 percent to 70 percent of the tested students, but that percentage depends on the test. Boer clarified that all students would be counted as part of a teacher's evaluation; those students would be counted, but the system described in Concept 1 would be sensitive enough to identify only if teachers are effective or ineffective at the outer edges. Concept 1 counts the students in the "not sure" category in the teacher's favor, giving the teacher the benefit of the doubt. After member questions, Boer and Dr. Ponisciak stated that PEAC would be responsible for creating business rules, cut points, and percentages.

A member raised the idea of using multiple years of student data, and Boer stated that using multiple years of student data would be possible under Concept 1 and would be one of the business rules to consider.

A member raised several concerns about district capacity to do this work and also provide teachers with information needed to improve student achievement. Boer stated that one of the benefits of using these national tests was the analysis done for districts. When pressed about whether national measures should be a smaller portion of the evaluation system, Boer disagreed, saying that when a teacher is not showing student growth, using a system such as the one in Concept 1 (which gives a teacher the benefit of the doubt) is important and should be a significant part of the teacher's evaluation.

Concept 2. Dr. Ponisciak explained Concept 2, in which the work is done at the state level. Using state-level data, a target of expected growth would be created for each student on the basis of the historical performance of similar students.

A member asked about how the state would differentiate teacher effects from peer/school effects because demographics are not the same in all districts. Dr. Ponisciak explained that the peer-effect issue needs to be explored. He raised the point about an "error of commission" rather than an "error of omission"; if the system makes an adjustment for a school or a class, it is possible to overdo it and create a bias in the other direction. Boer stated that the purpose was to keep the model simple. No methodology is perfect, but that is why there are multiple measures. Other measures, including student learning objectives and practice, also have measurement error; issues of demographics, as well as those issues, are not being controlled for in these areas.

A member asked about the weighting of the two growth measures. Boer clarified that PEAC already had talked about splitting the growth evenly between the two assessments. PEAC would need to discuss additional areas, including a classification system (determining how much growth should be considered *excellent, proficient, needing improvement, or unsatisfactory*).

After some questions, Boer explained that *growth* is defined as the difference between what a similar group of students would have done and what this student did. Stanton clarified that in Concept 2, the state is capturing and using more information, allowing it to be more nuanced in its conclusions. Because of the complexity of this model, most districts would not be able to do this nuancing on their own and would need state help. Boer said that there are a lot of good questions to be answered, and there must be a state model for Concept 2.

A member brought up the issue of district and state capacity and recommended assessing that capacity. Dr. Tomlinson stated that ISBE is working on an No Child Left Behind waiver right now and agrees that there must be multiple measures. She would like to talk with senior staff at ISBE about their capacity. Dr. Tomlinson then addressed some questions about what is currently in the Illinois waiver.

When two members stated that districts would have better capacity to implement Type II assessments, Boer questioned district capacity to develop valid and reliable Type II assessments, stating that if a district does not have the capacity to implement Type I assessments, its capacity to produce Type II assessments with those statistical characteristics seems unlikely.

Dr. Rutledge stated that although ISBE would look into state capacity, PEAC members should think about district capacity. The topic will be continued at the next meeting.

VII. Public Comment





Amy Alsop (IFT) expressed her concern about the discussion of a state model for student growth. She stated that she does not see how this system will help teachers improve instruction. She indicated that at least Type II assessments are closer to the classroom. Alsop also stated that the IFT did not come to the table to talk about sorting teachers and determining acceptable levels of growth. She suggested PEAC consider Type II assessments.

Kathy Shaevel (IFT) stated that schools should not use Type I assessments for all teachers schoolwide; such assessments should apply only to the teachers in that subject area. She stated that the impact that a drivers education teacher has on a student's reading score is different from the impact that the ELA teacher will have. She recommended that PEAC have a conversation about advice or guidance to the field about that issue.

Mary McDonald (CEC) announced the launch last week of the *Illinois Teacher Evaluation and Development* website, available at <http://ited.cecillinois.org>. Eighty districts signed up to access the site. CEC had student growth meetings with about 600 attendees.

VIII. Next Steps and Adjournment

Stanton summarized the next steps. ISBE will review its capacity and the way it has reacted to these proposals. Members will think about district capacity. The conversation on the guidance relating to the evaluation of special education teachers will occur at the next meeting. Members should be sure to link their organizations with the PEAC guidance online.

SYMBOL KEY	
	General guidance to all districts (includes topics related to principal evaluation, recalibration of evaluators, and so forth.)
	Required element of the state teacher evaluation model; any district unable to reach a collective bargaining agreement would default to these required elements of the state model.
	Recommended element of the state teacher evaluation model; all districts can use these elements as guidance and can adopt, adapt, modify, or ignore them as best fits their district needs.
	Communications document (shorter overview document) from PEAC to local education agencies.