Performance Evaluation Advisory Council (PEAC)

November 9, 2012

Meeting Minutes
Illinois State University Bone Student Center, Normal, Illinois

Members Present

Dr. Bette Bergeron, Benjamin Boer, Dawn Conway, Dr. Gail Fahey, Dr. Steven Isoye, Joseph Matula, Ernest Morgan and Dr. Yang Caroline Wang (representing Dr. Stephen Ponisciak), Janet Tate (representing Audry Soglin), Dr. Vicki Phillips, Dr. Diane Rutledge, Jodi Scott, Donald Daily, Richard Spohr, Dr. Gary Zabilka (representing Dr. Richard Voltz), Amy Alsop (representing Kristen Adams), Kurt Hilgendorf, Dr. Randy Davis, Stephanie Bernoteit, Michelle Standridge, Angela Chamness

Observers Present

 Bradford White (Illinois Education Research Council [IERC]); Jennifer Cowhy and Dr. Susan Sporte (Consortium on Chicago School Research [CCSR]); Kathy Shaevel (Illinois Federation of Teachers [IFT]); Christi Chadwick (Illinois State Board of Education [ISBE]); Gretchen Weber, Larry Stanton, Dr. Ellen Behrstock-Sherratt, Arie van der Ploeg, Nick Pinchok, and Jenni Fetters (American Institutes for Research [AIR])

Meeting Called to Order

The meeting was called to order at 10:13 a.m. by Dr. Diane Rutledge.

I. Welcome, Announcements, and Updates

Training Updates From Growth Through Learning: Mary Jane Morris, executive director of the Consortium for Educational Change (CEC), provided PEAC members with a written update on the evaluator training program. Dr. Rutledge noted that training videos are now available for use by teachers in individual or group settings. The link is available in the [November 9, 2012, *Newsline*](#).

II. Approval of the October Minutes

Approval of the October minutes has been delayed until the next meeting, on December 14, 2012.

III. Early Implementers Presentation—IERC and CCSR

Bradford White, Jennifer Cowhy, and Dr. Susan Sporte provided the Council with a brief presentation summarizing the key findings from their recent study, [*Designing and Implementing the Next Generation of Teacher Evaluation Systems: Lessons Learned From Case Studies in Five Illinois Districts*](#). The study collected key lessons learned from five Illinois districts that already have moved ahead with designing and implementing teacher evaluation systems. The research team conducted interviews with four to six key decision makers involved in the design and implementation of the system in each district. The study identified the following four challenges across the cases:

- Securing buy-in and understanding
- Using observations to improve instruction
- Building the capacity of educators
- Incorporating student growth

During the presentation, the authors shared the multiple strategies that the districts had identified to help address these challenges (see PowerPoint slides for details).

Council members had several questions, and the authors provided answers:

1. **Did the districts receive support or guidance from ISBE?** No, each district worked with either the Committee for Economic Development (CED) or an external technical assistance provider.

2. **What strategies worked best to improve communication and trust?** Several districts found that once they were able to agree on the general principles for the process, they were able to push aside past distrust and move forward.

3. **How did the districts make the decision to move ahead as an earlier implementer—was it a district decision?** The cases varied, but several of them were teacher-driven incentives. Dr. Rutledge noted that the professional associations already were working on how to support districts in ways to further coming together for these types of conversations, especially in districts with challenging climates. The presenters noted that part of the purpose of the study was to help other districts learn from the experiences of these early implementation districts.

4. **Do we have a sense of how many teacher leaders are participating as evaluators?** Dr. Rutledge was unsure but agreed to work with the CEC to try to find out, on the basis of the Growth Through Learning training.

IV. Communications Discussion

Dr. Ellen Behrstock-Sherratt provided an overview of the process for developing a communications plan. Council members had been asked to prepare for the November 9 meeting by connecting with their stakeholders and identifying key areas of miscommunication. Members reported the following critical areas that the Council needs to clarify:

- Implementation timelines and requirements—Some districts believe that there is a default model for teacher practice; some districts are limiting union representation and bargaining around student growth to the Performance Evaluation Reform Act (PERA) joint committee. Some districts believe that there is a single, joint committee for PERA and Senate Bill 7 (SB7).

- Whether or not the state model is a best practices model.

- Teacher evaluation for special education, gifted, and English language learners, especially around student growth.

- What is the purpose of the new evaluation system, other than it was created as part of a legislative effort to receive federal funds that the state ultimately did not receive? Many teachers believe that this system is being imposed without their input and without a strong justification.

- Some teachers are being told there is a preset distribution or "curve" that districts are expected to meet; other teachers are being told that everyone should be "proficient" or that there will no longer be "excellent" teachers in Illinois because the highest level in Charlotte Danielson's *Framework for Teaching* is so rigorous.

- Teachers often believe that the new evaluation system is about "fixing" teachers rather than recognizing that teachers always want to grow and improve. It is important to highlight and value the formative aspects of the evaluation process.

- Teachers are unsure about the training process for evaluators and what that process

entails.

- Whether or not districts can or should modify the Danielson *Framework* and still be in compliance with state requirements for a research-based framework.

- How do the new evaluation systems align with Common Core implementation?—Other states have demonstrated how the Danielson *Framework* aligns with Common Core requirements.

Dr. Behrstock-Sherratt asked Council members to reflect on why they believe that the evaluation reforms had been undertaken and what their purpose should be. Council members offered a number of perspectives, including the following:

- PERA is a transition to a new framework to evaluate teachers fairly, and it accurately differentiates teaching performance.

- PERA is not about labeling teachers. Its purpose is to create a better system for teachers to receive feedback from evaluators by creating a roadmap or guide for a vision of what good teacher looks like. An instructional practice rubric empowers a teacher to see where improvement is needed and provides an anchor for the conversation with the evaluator.

Council members shared lessons learned from previous reform efforts, including the following:

- Current timelines are creating a very rushed implementation process. It would be better if key decisions, guidance, and piloting could be completed before districts need to implement.

- The public forums held last year by PEAC allowed ISBE to hear directly from teachers and enabled teachers to voice their concerns. The information they shared was used in our PEAC deliberations.

- In the past, the framing of how important teachers and leaders are for students and their learning has not been as focused as it could be. The goal should be framed as helping teachers to improve so that their students can improve. PEAC needs to distill this message so it can be used continuously.

- Actions speak louder than words. Communicating the right message is important; but if our actions are the opposite of what we say, the message is useless. For example, we told administrators that we value their time; however, during the evaluator training process, administrators believed that they did not have adequate time to complete the training in a meaningful manner. We cannot reinvent history—the legislation will affect teachers' jobs. We have to acknowledge that conclusion openly and encourage everyone to see this situation as an opportunity to make the best we can of something difficult.

- The PERA legislation raises a larger issue: saving public education itself. PERA is part of a larger push for recognizing the fundamental, foundational importance of public education; teachers and teacher leadership are part of that foundation. We need a message that unifies and helps all of us buy in to the best version of ourselves possible.

Dr. Behrstock-Sherratt asked Council members to identify the dissemination channels that would be the most effective for communicating with the field. One member noted that whichever channels were selected, it is important to pay equal attention to getting comments,

feedback, and input from the field in order to make communication a conversation. Another member noted that it will be important to coordinate across professional associations to ensure that the same, consistent message is being communicated. Dr. Rutledge noted the importance of clarifying for stakeholders how any feedback or input from the field is actually used and how it influences the process. Another member emphasized that communication channels should highlight the ongoing importance of teacher collaboration in implementing the evaluation process. As a final note, a council member reminded the Council that PEAC's role is advisory and that ultimately ISBE owns the messaging and communication.

Next Steps: AIR will use information from the discussion to draft a communications plan to share with the Council at the December 14 meeting.

V. Topics in Stage 2: Consider

> **SM: Require**

> **SM: Recommend**

A. Student Growth Model for Teachers

Nick Pinchok provided an overview of process used by other states for creating, selecting, and recommending assessments to districts for use in teacher evaluation, and he highlighted examples from New York, Massachusetts, and Chicago. Examples of the review criteria used in each example were circulated to the Council for consideration.

Arie van der Ploeg gave a presentation describing the key differences among 11 Type I assessments that PEAC could consider. He provided a brief overview of each assessment, why it was created, its strengths and weaknesses, and its specific purpose. He then outlined the key choices that the Council will need to make in selecting assessments and suggested that the Council consider the trade-offs that will be required regardless of which assessments are selected. Specifically, the Council should consider the following:

- Construct validity, relevance, and utility
- Type of comparison method (e.g., absolute versus relative)
- Metric, scale, and accuracy
- Item format, delivery method, scoring, and reporting formats

Finally, van der Ploeg noted that demonstrating student growth on an assessment necessarily requires demonstrating change over time on a scale(s) of status. In order to be reliable, value-added or student-growth percentile models require large numbers of students and teachers; these approaches to growth may not be appropriate in smaller districts.

Ernest Morgan, Dr. Yang Caroline Wang, and Dr. Stephen Ponisciak (by phone) briefly discussed a study they are completing. The study examines how well the Measures of Academic Progress (MAP) assessment and SAT 10 functioned, using a value-added measure of student growth. The team hopes to share the completed study with PEAC in December.

Pinchok provided an overview of how four school districts in Illinois are addressing student growth: Niles Township High School in District 219, Plano School in District 88 (K–12), Rockford Community Unit School in District 205, and Wilmette School District 39. Pinchok presented an updated Summary Table

of Assessment Characteristics and proposed a list of administration criteria to consider.

Council members noted that Niles Township High School's growth model has been very expensive to create and implement. The district has had to hire a full-time data analyst to process the data, and some teachers in the district do not have access to the data.

Another member asked if any of the Classroom Assessment Techniques (CAT) assessments could be given in a paper-and-pencil format. The presenters noted that several of the assessments are offered in paper format; however, it takes the students significantly longer to complete the assessment process.

One member highlighted a comment by van der Ploeg that because assessment companies are marketing their products, it is important to verify carefully all the claims made (e.g., Common Core alignment). One member stated that she preferred the Chicago Public Schools Request for Qualifications (CPS RFQ) rubric that provided broad categories, with details underneath to help with selection, and she suggested that this rubric could be a very useful format for districts in developing PERA guidance about selecting assessments. Districts are likely to be much more comfortable if they were presented with a narrow menu of options and allowed to select the right assessments for their schools.

Pinchok reminded the Council that the state default model was a set of requirements that a district unable to reach agreement must default to. In these rare cases, the state model must clearly identify what assessments the district must use.

Two members suggested that it was not necessary for the state default model to mandate a single assessment for each teacher category; rather, the state model could provide a narrow list of perhaps three options that a district must select from and then complete a Request for Proposal (RFP) process in order to make the final selection.

Dr. Rutledge noted that the original purpose for identifying a single assessment was to enable a statewide bidding process for reduced rates.

Another member raised a critical concern about the overall cost to calculate growth scores and develop the necessary analytics and infrastructure. Many districts have developed local item banks, but they lack the analytics needed to calculate a student growth score. It is important not to build a set of recommendations that districts cannot afford. It also is important that student data are not collected without investing in the human capital and foundational knowledge that teachers and administrators need in order to put the data to good use. PEAC must think about best practice first and foremost in making these decisions.

One member noted that the weight of this decision was significant and reiterated the importance of not rushing to a decision, given the impact it will have on districts.

Members asked why the Partnership for Assessment of Readiness of Colleges and Careers (PARCC) assessments were not on the list for consideration. Larry

Stanton noted that the PARCC assessments were still under development and would not be available until 2014.

Members asked that AIR develop a menu of six or seven assessments, presented with in-depth information (such as the CPS review rubric), which could be used by districts (defaulting or not) to select assessments. The more information and flexibility the Council can build in to the system, the less pushback the move to student growth measures is likely to generate.

Next Steps: AIR will bring a draft of several assessment menu options and a rubric similar to the CPS RFQ rubric to the December 14 meeting.

> **Guidance**

B. Growth for First-Year Principals and Assistant Principals

Jenni Fetters provided an update to the Council on efforts to reach out to districts and gather information about the way that districts were measuring student growth for first-year principals and assistant principals; thus far, only one district has been willing to share that information. To try to reach more districts, Fetters noted that several professional association representatives on the Council had suggested that a request for this information be included in the Superintendent's Weekly Message. She provided the Council with a draft of the questions that would be included in the request.

One member noted that districts had expressed their concern that if they had shared their process but had not yet decided how to measure student growth for these categories of staff, they did not want to be identified as noncompliant. Fetters agreed to rework the information request to ensure that the districts would know that the information would be collected and made anonymous before being shared with PEAC or ISBE. Council members also suggested adding a question that asked, "If your district has not made a decision about this topic, what key barriers or challenges have you encountered, and what supports do you need to move forward?" Another council member noted that the Compliance Division was requiring districts to submit information on six dimensions, one of which explicitly asked about student growth for first-year principals.

Next Steps: AIR will revise the request for information accordingly and ensure that it is included in the next Superintendent's Weekly message. At the next meeting, all information gathered will be summarized and used to create a draft guidance document for PEAC to consider.

VII. Topics in Stage 2: Consider (Cont.)

> **SM: Require**

> **SM: Recommend**

C. Operating Guidelines for Student Growth

Gretchen Weber guided the Council through a discussion of proposed operating guidelines for student growth. The Council was presented with each of the following questions and proposed answers. The proposed answers were provided for Council members to react to and discuss and do not constitute recommendations or decisions by either AIR or PEAC. *Note:* The order of the questions reflects the actual sequence of discussion; therefore, the question numbers are not sequential but do align with the question numbers on the

accompanying Powerpoint presentation.

Question 1: Which students are counted for a teacher's growth measure?

Council members raised concerns about how this rule should be defined in (1) instances where a student moves midyear from one teacher to another teacher, and (2) in response to intervention (RTI) situations. Council members also wanted to know what types of roster verification processes would be provided to teachers to ensure that the students associated with their names were the correct students and had been in their courses an appropriate length of time. Representatives from the Value-Added Research Center (VARC) noted that a student could be attributed to more than one teacher; however, weight of the student's score must be divided appropriately among all attributed teachers.

Question 4: How do we attribute growth to teachers who share responsibility for students?

One council member wanted to know what portion of the student's growth would be assigned to a limited or shared category of teachers. Another member raised a concern that dividing the teachers by level of responsibility may make the system too complicated, and another member noted that CEC does not recommend apportioning student growth by level of responsibility. He stated that he prefers the joint responsibility approach instead. One Council member, however, believed that if the system did not differentiate levels of responsibility, it might be more difficult to convince teachers to accept student teachers. The Council members requested that AIR investigate instances where these two different approaches have been implemented and discover whether any of these concerns were valid or had presented problems in other states.

Question 5: How are later teacher assignment, absence, and transfer accounted for in student growth measures?

One Council member noted that the proposed answer seemed very complicated and stated that she preferred the earlier answer that only students who were in the class for the pretest, posttest, and most of time in between should be counted. Another member cited a research study by CCSR that found that for every five days a student misses a class, there is a corresponding 15 percent drop in the likelihood that the student will stay on track for graduation. In addition, district policies on discipline and the extent to which students are removed from the classroom for discipline also affect how a district might answer this question. One Council member was very concerned about the management and precision of the data required for these types of very specific rules. How would the necessary data infrastructure be implemented, instituted, and maintained?

Stanton reminded the Council that 80 percent of students would be addressed by the pretest and posttest rule; the focus of this question was about the remaining 20 percent. Another member was concerned about the legality of the requirement to exclude teachers with fewer than 10 students from being assigned a growth rating—the legislation requires a growth rating for all teachers. VARC representatives noted that this rule was a statistical necessity—with fewer than 10 data points, a growth measure's reliability would be in question. Another

member suggested that these types of problems (e.g., where the legislation is in conflict with a statistical impossibility) be left up to professional judgment. Stanton noted that one possibility might be to attribute the school's average student growth to these teachers. Several members agreed that this approach might be a good one because it would emphasize teacher teams and collaboration; however, several other members raised concerns that attribution on the basis of team responsibility can be very difficult. It is important to remember that evaluations deal primarily with individuals. Although encouraging teamwork and collaboration is important, it is equally important to remember that teachers are still individuals.

Question 7: What happens when assessment data are missing for a student or group of students?

Council members largely agreed that a student without a pretest or posttest score could not be counted in a teacher's student growth rating. One Council member wondered how long after a pretest had been given should a student be allowed to take a make-up test. Another member was opposed to allowing a pretest score from another school (in the case of a transfer) to be counted; even if the same assessment was used, the same protocols (length of time allowed, and so forth) may not have been followed, and, therefore, in that member's estimation, the two tests were not comparable.

Question 10: Should the state model require the use of multiple years of assessment results to evaluate student growth?

One Council member wanted to know the implications of multiyear data for fairness and accuracy. VARC representatives noted that generally speaking, the more years of data, the better the data's accuracy; however, it is important to be careful when comparing teachers on different time periods in the school.

Question 2: How long does a student need to be in a teacher's class for the teacher to contribute to his or her growth?

Many Council members believed that the attendance cut-off should be 95 percent, in order to align with the state's truancy requirements. Members noted that the rules would need to address students receiving homebound instruction or students who are absent for college visits or workshops because these students will still be on a teacher's roster. The distinction between counting students as being in the classroom versus as being in attendance was raised as problematic (e.g., the proposed answer currently says *in class* for 60 percent of the time). Several members noted that truancy requirements specify "unexcused" absences. Several members preferred attendance to be used as the determinant because it is already recorded. Several members believed strongly that the percentage should be based on whether students were recorded as in class and *not* on whether they were counted for attendance-records purposes, noting that teachers cannot influence student learning if the students are in the nurse's office during that period (e.g., they are still in attendance but not in class). Council members requested that AIR gather information on the percentage cut-offs that other states have adopted.

Question 3: What portion of a teacher's students need to be included in the

growth measure?

Members believed that all students for whom a measure was available should be included.

Question 6: How should the state model ensure that student rosters are accurate and the correct students are being included in a teacher's growth measure?

Council members agreed that the state should provide guidance on what a good roster verification process entails and what kinds of mechanisms might be available for teachers, not just districts, to verify a roster.

Next Steps: AIR will use the discussion to craft a draft guidance document for PEAC members to review and discuss at the December 14 meeting.

VIII. Topics in Stage 3: Decide

> SM: Require

A. Summative Rating Definitions

Stanton guided Council members through a series of proposed summative rating matrixes and a process for addressing major inconsistencies between teacher practice and student growth. The presentation was based on the Council's discussion on October 19 and on additional matrixes submitted by members in the interim. In addition, Stanton provided feedback information provided by ISBE's General Counsel regarding the legality of using a matrix approach to satisfy PERA requirements that student growth scores constitute 50 percent of the final summative score. According to the General Counsel, the matrix approach does meet the requirements of PERA, with one exception. The proposal to require additional evidence for a two-level inconsistency (see the green boxes in the proposed matrix) was rejected by the General Counsel because the middle rating (between the two levels) is considered a clear resolution that maintains a 50/50 balance between the practice and growth scores.

Stanton also provided PEAC members with a version of the matrix that aligns with the matrix proposed for the principal evaluation model. Stanton noted that the principal matrix does not use a consistent criteria for selecting ratings for one-level and two-level discrepancies. According to the General Counsel, this would not preserve the compliance with the 50 percent rule.

PEAC members raised several questions and concerns. Members wanted to ensure that the column labels and row labels on the teacher matrix and principal matrix were consistent with each other. Several members strongly questioned the reasoning behind automatically selecting the rating between the two scores when there is a two-level discrepancy (e.g., the green boxes in the proposed matrix). Members believed that to simply say a teacher's score was the average of the two levels without requiring additional evidence or at least a reexamination of the evidence was a rather significant assumption. It was also not clear why requiring additional evidence would make the decision rule inconsistent with a 50/50 approach. Stanton agreed to raise these questions with the General Counsel and to seek additional clarification for the December meeting. One member requested that additional information be sought on what other states are doing related to

using matrixes with a 50/50 rule. The General Counsel is providing only a legal opinion, and PEAC must avoid the expensive situation where the state model cannot stand up to legal challenge. A member asked what was the process for arriving at the professional practice rating and the student growth rating. Stanton noted that this issue, although very important, required a separate discussion. Members requested that the process for resolving three-level discrepancies use the term *qualified evaluator* rather than the term *district administrator,* specifically because the definition of *qualified evaluator* was provided in SB7.

Next Steps: AIR will seek additional clarification from ISBE's General Counsel on the two-level descrepancies.

IX. Topics in Stage 4: Give Guidance

Due to time constraints, the Council's discussion of the Student Learning Objective (SLO) Guidance materials and the teacher evaluation nonobservable components was delayed until the December 14 meeting. Council members were asked to send all comments on these documents by November 30, 2012, and to contact Fetters if they are interested in contributing to the creation of Illinois-specific SLO examples for the SLO Guidebook.

X. Public Comment

Kathy Shaevel requested that page 3 of the Guidance on nonobservablecomponents be revised for clarity. The document suggests incorporating artifact review in the preobservation and postobservation conferences; however, for the summative conference, the artifacts would need to have already been received in order for a summative score to be determined. The discrepancy needs to be addressed.

<div style="border:1px solid black">

SYMBOL KEY

</div>