

# CAEP Standard 4 Evidence: A Resource for EPPs

Addressing components:

- 4.1 | P-12 student learning and development
- 4.2 | Teacher observation evaluations and/or student perception surveys

Contents:

- Challenges of Standard 4 | Page 1
- Purpose of Standard 4 | Page 2
- Evidence provided by EPPs | Page 2
- Examples from self-study reports | Page 4

## THE CHALLENGES OF STANDARD 4

Standard 4 is challenging for educator preparation providers (EPPs). It is intended to provide evidence of preparation results when completers have been employed by schools—where preparation matters most. These inservice measures are a new category of evidence for most EPPs, are often difficult to obtain and are sometimes withheld due to interpretations of federal privacy protection laws.

There is a suite of measures, tracking through the four components of Standard 4, which are meant to provide multiple perspectives of completer efficacy on the job.

1. One explicitly describes the influence that teachers have over their P-12 students' learning and development.
2. Another is a classroom performance measure taken either as a structured observation evaluation or in the form of student perception surveys on their classroom experiences.
3. The third is employer satisfaction, including satisfaction as reflected in teacher retention and employment milestones (the next level of certification, for example, or promotion along a new teacher's career path).
4. The final measure is completer satisfaction with their preparation—do they believe, after their experience on the job, that they were appropriately prepared for the daily responsibilities they have encountered.

Acknowledging the challenge of attaining appropriate measures for each of these critical topics, a focus group participant in a CAEPCon session recently observed, "Standard 4 asks the right questions," and that is the critical point. The answers to these questions provide multiple perspectives on teacher readiness, and for that reason the combination is a summative compilation of information on the effectiveness of educator preparation, as well as a powerful source of information for EPP improvement and monitoring of success.

## THE PURPOSE OF STANDARD 4

Standard 4 serves a unique purpose among the five CAEP standards. It is part of a set recommended by the Commission on Standards and Performance Reporting in 2013. The first three standards address the principal aspects of the preparation experience:

- Standard 1—candidate content and pedagogical knowledge and the ability to apply them;
- Standard 2—clinical experiences and partnerships; and
- Standard 3—monitoring proficiencies of candidates as a continuing responsibility.

To complement the first three standards', focus on preparation experiences, Standard 4 addresses results. It asks about effectiveness of completers when they are on the job. Adapting ideas from the Baldrige Education Criteria for Excellence, Standard 4 is CAEP's way to indicate that results matter and that gathering and systematically reviewing data about results is an essential element of any EPP's responsibilities. The Standard 4 measures make up half of a set of eight annual reporting measures that constitute a "dashboard" of key preparation indicators. EPPs monitor data from these indicators as a means to evaluate their progress, inform any needed adjustments, and they post the data on their website for purposes of accountability to stakeholders and of informing potential candidates. The remaining four indicators describe outcomes, including licensure, employment, completion and consumer information (such as typical employment following preparation, or cost of attendance for preparation).

## EVIDENCE PROVIDED BY EPPs

EPPs have followed the CAEP Standards and Handbook procedures, and the CAEP review procedures have been in place with site teams and the Accreditation Council for decisions in October 2016 and April 2017. So now, for the first time, CAEP is able to describe what EPPs are including in their self-study reports as evidence and to provide actual examples.

The EPP self-study report evidence can be categorized by type. Drawing on the 17 cases decided in April 2017, we find that about one third of the evidence was for State or district measures of P-12 student learning or growth, and another fifth were case studies to create student learning data. One EPP provided student survey results, and another offered a teacher evaluation tool as evidence.

All of these items are direct measures of what P-12 students have learned or of teacher performance in the classroom. Data sources for EPPs include information shared from states, case studies devised by the EPP, information available from local school districts where teachers are employed, and data from state or district teacher evaluations or student perception surveys. EPPs are supplementing state data with case studies that will provide complementary information about other grades or subjects not included in state data. They are also taking advantage of unique data available in districts where their completers are employed—from NWEA, for example, or ACT Aspire. Evidence of this type accounts for about 60% of what EPPs submitted.

The EPP self-study reports also make clear that the source of 4.2 evidence and 4.1 evidence is frequently the same—that is, state conducted teacher evaluations often combine some measures of P-12 student learning together with observations of teacher performance in the classroom.

An additional finding was that some evidence was inconsistent with Standard 4. For example, some EPPs tried to use candidate data and survey opinion data not clearly related to measures for the 4.1 or 4.2 components. In other cases, plans were submitted as evidence—which CAEP had encouraged for new forms of evidence—but there was no documentation that steps outlined in those plans had actually been executed at the time of the site visit. There was little or no observable progress toward the needed evidence. Evidence falling into this category accounted for about 40% of what was submitted. As EPPs become more familiar with the CAEP Standards and with guidelines for self-study reports, this off-the-mark evidence should decline and disappear.

# Examples for component 4.1, P-12 student learning and component 4.2, teacher observation evaluations and/or student perception surveys

*P. 5: Example #1—STATE University—P-12 academic achievement comparison using NWEA data, with confirmation from correlated measures and case study using Danielson.*

Makes use of data in the state from NWEA-available student growth measures; case study includes TWS type data that is linked with similar assessments used during preservice; cites findings from research that associate teaching strategies that demonstrate an impact on student learning; mentions importance, first, of building a tracking system for graduates.

*P. 9: Example #2—PRIVATE University—P-12 student growth complemented by planned teacher action research*

Self-study report includes student growth percentiles from state but the information is highly summarized, so of little utility to the EPP. Currently piloting a teacher action research project using volunteering completers, and being constructed as a continuing activity. The design permits links with pre-service data for the same completers. Candidate tasks are similar to those in a teacher work sample assessment, and specifically include pre- and post-measures for teaching a comprehensive unit.

*P. 13: Example #3—PUBLIC University—P-12 student value-added data as part of state teacher evaluation, complemented by planned teacher action research*

State value-added data are available but are aggregated limiting utility. But these data are part of the state's teacher evaluation assessment so evidence for 4.1 and 4.2 are linked. Characteristics of the state system are described. The EPP shows options that it has considered to complement the state data, and designates one as the path it intends to follow. It will work with a school partner to gather and evaluate classroom data from novice teachers.

## EXAMPLE # 1: STATE University—P-12 academic achievement comparison using available data, with confirmation from correlated measures

IHE enrollment 23000; EPP enrollment around 2300 undergraduate and graduate

### PART 1: P-12 STUDENT LEARNING FROM A RESEARCH STUDY USING AN EXISTING DATA SET

NOTE: This part of the STATE University evidence was taken from a research study that includes STATE University completers and also other completers from the same state. The numbers are only a small proportion of the EPPs completers. It would be useful if the self-study report had been more explicit about the sampled participants in relation to STATE's completers in terms of demographics or employment sites.

Research study population: graduates of STATE university in 2011-2014 (n=15) and all teachers in state who graduated from universities in the state from 1991-2014 (n=82).

Comparisons: All STATE university with other providers; 2011-2014 STATE university graduates with those of other providers

Study population: from 8 schools, teaching in grades 1-5

Study subjects: math and reading

Study assessment: Nationally normed, NWEA: Measures of Academic Progress (MAP)

EPP provides key information about this data set that is used as partial evidence for 4.1

Take advantage of available data. NWEA emphasizes growth measures and uses computer adaptive tests. Some 8500 schools, districts and states participate.

Year 1 includes the average end of year percentile rank of the students of completers after receiving professional development, and year 2 is one year after.

**Table 1. Average End of Year Percentile Rank of Completers by Group and Year**

Group	Mathematics	Reading
	Year 1	Year 1
STATE University	63.8 (1.3)	62.0 (1.1)
Other providers	62.0 (1.2)	62.3 (1.0)
	Year 2	Year 2
STATE University	64.9 (1.0)	62.7 (1.0)
Other providers	62.7 (2.9)	61.6 (2.8)

EPP provides actual data—in this case, comparison of STATE university completers with other providers in the state and later in terms of national norms; contextual information, particularly for similar EPPs, completers, and students is useful.

Note: Number in parentheses is the standard error of the mean

EPP provides information on statistical tests and results—a data quality point

In Year 1, there were no differences across groups on student achievement for mathematics [ $t(80) = .34, p = .73$ ] or reading [ $t(80) = .55, p = .58$ ], but both groups were above the national norms for the test of mathematics and reading (i.e., 50%). Likewise, in Year 2, there were no differences across groups on student achievement for mathematics [ $t(102) = .74, p = .46$ ] or reading [ $t(102) = .37, p = .71$ ], but both groups were above the national norms for the test of mathematics and reading.

**Table 2. Average Change in Percentile Rank of Completers by Group and Year**

Group	Mathematics	Reading
Year 1		
STATE University	6.7 (1.1)	2.6 (0.8)
Other providers	6.0 (1.1)	2.4 (0.7)
Year 2		
STATE University	9.6 (0.8)	5.5 (0.6)
Other providers	7.1 (0.9)	4.0 (0.6)

These are data on change between year 1 and year 2

Note: Number in parentheses is the standard error of the mean

For Year 1, change in performance from the beginning of the year to the end of the year did not differ significantly across groups for mathematics [ $t(80) = .49, p = .63$ ] or reading [ $t(80) = .25, p = .80$ ]. However, for Year 2, change in performance from the beginning of the year to the end of the year was significantly greater for STATE University completers than for other completers for mathematics [ $t(102) = .2.1, p = .04$ ] and marginally greater for reading [ $t(102) = 1.7, p = .09$ ].

These data are for 2011-2014 completers of from STATE University compared with 2011-2014 completers from other universities in the state.

Student achievement was also compared for more recent completers (N = 15).

**Table 3. Average End-of-Year Percentile Rank of Recent Completers by Group and Year**

Group	Mathematics	Reading
Year 1		
STATE University	66.8 (4.2)	66.4 (1.8)
Other providers	61.3 (2.7)	60.6 (2.5)
Year 2		
STATE University	66.8 (4.2)	66.4 (1.8)
Other providers	61.3 (2.7)	60.6 (2.5)

Note: Number in parentheses is the standard error of the mean

“No difference” may be attributable to the small sample.

There were no differences across groups on student achievement for mathematics [ $t(10) = 1.1, p = .28$ ] or reading [ $t(10) = 1.8, p = .09$ ], but both groups were above the national norms for the test of mathematics and reading.

**Table 4. Average Change in Percentile Rank of Recent Completers by Group**

Group	(Year 1)	
	Mathematics	Reading
STATE University	7.1 (2.4)	3.3 (1.7)
Other providers	8.3 (3.6)	4.6 (1.5)

Note: Number in parentheses is the standard error of the mean

Change in performance from the previous year did not differ significantly across groups for mathematics [ $t(10) = .25, p = .81$ ] or reading [ $t(10) = .58, p = .57$ ].

Because the NWEA data cover only some subjects and grades and may not be available again, the EPP undertook additional steps to provide evidence for 4.1

## PART 2: OBSERVATIONS AND CORRELATED MEASURES:

Study (as described by EPP):

- 13 teachers, grades K-6, in multiple school districts, within the EPP's state. All participants were in their first two years of teaching.
- Each completer taught three language arts lessons and three math lessons across the Spring 2015 semester. One lesson per completer was digitally recorded.

The EPP should provide specific information about the representativeness of sample.

The EPP did not break out these data with regard to what was found about student learning. The approach is similar to Teacher Work Study approaches.

Six classroom observations were conducted by [Danielson Framework for Teaching](#) certified observers using observational field notes and the Danielson FFT Evaluation. Additionally, completers participated in a statewide alumni survey and participated in two focus groups that were recorded and transcribed verbatim. To further strengthen links between teacher practice and the teacher education program, additional data sources include

Note: EPP links to other data with the intent to learn more from the case study.

- (1) Preservice professional year evaluations;
- (2) Inservice teacher evaluations using the FFT;
- (3) Employer interviews; and
- (4) Employer survey data.

Completer study observations were conducted using the observable components from the Danielson FFT: Domain 2 Classroom Learning Environments and Domain 3 Instruction.

Both domains were high overall:

- Domain 2: Learning Environment average of 2.81 and
- Domain 3: Instructional Practice average of 2.81.
- The highest area of strength (3.0) for completers was in the area of 2A developing respect and rapport. This is consistent with findings on the Tripod survey for the construct of Care. This means classroom interactions were observed to be warm and caring with respect shown for cultural and developmental differences among groups.
- Another completer strength included 3b questioning and discussion techniques. With an average of 2.91, 3b was highest in Domain 3 Instruction. Completers were strong in their ability to use discussion techniques with a high level of questioning. This is also consistent with findings on the Tripod survey of student perceptions.

Danielson components are evaluated against rubric descriptions for four levels of proficiency, each uniquely fitted to the focus of the component. A "1.0" would represent the lowest level, and "4.0" the highest. The individual component scores are averaged for the domain (e.g., learning environment) score.

EPP triangulated results from different sections of Danielson and from Tripod

EPP notes relationships in the data they have found

- Completers also demonstrated their ability to clearly and correctly communicate content knowledge through the use of academic vocabulary and articulate student learning targets with a high rating in 3A communication (2.85). Attention to academic vocabulary within the content area is critical for student success. Classrooms that directly teach academic language as part of the lesson focus, increase scores by as much as 30 percentile points (Flanigan, & Greenwood, 2007).

Additionally, observers looked for enactments of teaching strategies that have consistently demonstrated an impact on student learning in educational research. These data were collected using a check sheet of High Leverage Practices based on the research of Marzano, Pickering, & Pollock (2001); Depth of Knowledge (Bloom et al, 1956, Webb 1959); Engaging Qualities (Schlechty, 2002; Antonetti & Garver, 2015); and enactments of the College and Career Ready Standards based on the EPP’s state focus on the Common Core Standards shifts of instruction in Math and English Language Arts (CCSS Shifts, 2014).

EPP links teaching strategies to impact on student learning in addition to research cited in this paragraph, the EPP self-study report cites the MET Study (Gates Foundation, 2012).

**AN OBSERVATION FROM THE EPP:**

One of the first things STATE University faculty needed to purposefully implement the call for data in CAEP Standard 4 was to track placement of graduates.

Note where the EPP started—by tracking completers

- In Spring 2014, faculty began compiling spreadsheets of completer placement.
- By Spring 2015, the faculty had a high percentage of placements confirmed.
  - With 185 completers in 2014, employment status of 98 teachers was confirmed. Sixty-nine percent of these completers (68/98) were employed in a field for which they were prepared in their initial preparation program.
  - Sixteen percent were employed in an education position outside of their preparation.
  - Only seven candidates were not employed and 5% were pursuing higher education degrees.
- With renewed efforts at tracking completers, more than 69% should be confirmed in placements after graduation in future years. These numbers will support tracking important employment milestones and retention. Having employment confirmation also allowed for more purposeful contact of employers (i. e., direct name correspondence) and larger databases for alumni surveys

## EXAMPLE #2: PRIVATE University—P-12 student growth complemented by planned teacher action research

Private, religious affiliation, IHE enrollment around 3100; EPP enrollment around 80

EPP description of what the state provides and its utility to them

### PART 1: P-12 STUDENT LEARNING IN STATE TEACHER EVALUATIONS:

- State growth percentile model—assesses growth by comparing individual students to students who begin at similar levels of achievement to determine relative growth over time; teachers evaluated annually in one of four categories; every evaluation system must incorporate measures of student growth and achievement as a significant portion of a teacher’s evaluation
- The EPP does not have access to student growth associated for each completer—just the PRIVATE university composite. The state does provide PRIVATE university comparisons with other EPPs in the state.
- 99% of PRIVATE university completers score as “effective” or “highly effective”

State four-point scale is described as “ineffective,” “needs improvement,” “effective” “highly effective,” but is not further explained by the EPP in its self-study. The “comparisons with other EPPs in the state” are not provided in the self-study report but would help put the PRIVATE U. data into context.

First approach to complement limited state data for 4.1: Triangulate data from different sources.

### PART 2: OTHER MEASURES

Supporting or confirming evidence from these sources:

- Added questions to completer surveys in first, second and third years of teaching asking about the growth of their students. Voluntary, self-reported.
- Added question to employer survey about the impact of PRIVATE university completers on P-12 student growth—not for each individual teacher, but for composite across all teachers from PRIVATE university.
- At end of pre-service student teaching, ask completers to reflect on the effects of their actions—and adaptations made in their practice—to meet the needs of each P-12 learner. What they learned: candidates with longer placements (yearlong) outperformed those with only one placement in one semester.

The added questions would be part of a component 4.4 measure.

The employer survey addition would be part of the component 4.3 measure.

Building on the preservice TWS as a base point for 4.1 inservice evidence (see below).

Also, those with two placements during one semester of student teaching outperformed candidates with only one placement.

- The Teacher Work Sample produced during student teaching serves as a baseline for a pilot study for an eventual teacher action research case study.

Connecting evidence for 4.1 with data collected at the candidate exit, as part of component 3.5 evidence, provides additional context.

### PART 3: TEACHER ACTION RESEARCH CASE STUDY PILOT

- The lack of available evidence along with the lack of depth in analysis of the data that is available for Standards 4.1 and 4.2 from the state and schools has led to the development of a case study pilot for completers beginning with the academic year 2015-2016.
- Some additional outcomes of the case study may be to determine how to establish accurate inter-rater reliability and training on rubrics used for the assessment; to determine whether we need to align the Teacher Work Sample (during student teaching) and TAR (teacher action research) more closely with the state teacher evaluation model.
- Study uses a cohort of completers who are willing to participate in a presentation as they go into their first, second and third years of teaching.
- Will begin a new cohort each year with a minimum of 15% of completers participating.
- Foci are (a) instructional decision making; (b) analysis of student learning; (c) reflection and self-evaluation.
- Participants in the case study will develop a presentation on those foci and share the results of their action research with their cohort group. Participants will self-evaluate using the rubric. Cohort groups will follow a Critical Friends Protocol to assist their peers in reflection and to provide constructive feedback. Participants work will also be evaluated by the EPP and will be shared with the cohort group in the aggregate. This will provide a basis for their feedback to the EPP on areas of strength and need. Additionally, it will provide the EPP with a comparison following a completer from student teaching through the third year of teaching evaluation.
- 23 of 29 completers from a recent cohort of completers agreed.
- EPP assembling data on representativeness of potential participants, including their scores on the pre-service teacher work sample.
- The instrument measurements are derived from the Teacher Work Sample items relevant to student learning and teacher effectiveness, which are used as a baseline.
- In the years going forward, EPP created measures and data will be compared with new requirements of the state that will provide data to EPPs to determine relevance, reliability and validity of the case study. The EPP will also consider if there are redundancies that can be eliminated.

EPP recognizes limitations of available data and proactively explores other avenues for more specific/complementary data to strengthen case by providing more actionable data for improvement.

Participants are "willing completers"

Ideally, the EPP would also have commented on representativeness due to participation bias.

Linking in-service findings with pre-service findings for the same completers/teachers

When this evidence was submitted, plans were permissible as evidence. Note, too, that with this and other evidence, the EPP accumulates additional data over time, so faculty can be more confident in their analyses and interpretation of results.

EXCERPT from description of what participants in the action research will do:  
*Description of the tasks that participating teachers agree to conduct as action research*

<b>The Teacher Action Research Process</b> <i>Before beginning - read all rubrics for each section AND review the timeline.</i>	
<b>STEP 1 - Select a focus:</b> <input type="checkbox"/> Review Contextual Factors that affect your teaching and learning situation <input type="checkbox"/> Look at Student Learning Objectives (SLOs) to help determine areas of need <input type="checkbox"/> Determine your area of focus	<b>Timeline:</b> Completed BEFORE unit starts
<b>STEP 2 – Clarify theories</b> <input type="checkbox"/> Review research pertaining to your area of focus <input type="checkbox"/> Determine what you believe will make a difference to student learning	<b>Timeline:</b> Completed BEFORE unit starts
<b>STEP 3 - Identify research questions</b> <input type="checkbox"/> Define 1 to 3 research questions you wish to answer at the end of this action research project	<b>Timeline:</b> Completed BEFORE unit starts
<b>STEP 4 – Collecting data</b> <input type="checkbox"/> Develop a comprehensive unit plan that matches with your focus and research questions and includes lesson plans <input type="checkbox"/> Include an overview of your assessment plan <input type="checkbox"/> Create Pre-and post-assessments <div style="border: 1px solid black; padding: 2px; margin-top: 10px; width: fit-content;">           Source of data on P-12 student learning         </div>	<b>Timeline:</b> <input type="checkbox"/> Completed BEFORE unit starts, with adjustments made while teaching the unit based on student response and needs <input type="checkbox"/> Completed BEFORE unit starts, with adjustments made during and at end of unit
<b>STEP 5 - Analyzing data</b> <input type="checkbox"/> Analyze results of pre-assessment plan and make any adjustments to unit based on student needs <input type="checkbox"/> Generate and analyze Graphs/tables of student learning to make adjustments to unit as needed and to assess results at the end of unit <input type="checkbox"/> Summary of class learning (Post assessment data) <input type="checkbox"/> Summary of Subgroup differences <input type="checkbox"/> Reflection of Unit	<b>Timeline:</b> <input type="checkbox"/> Before unit starts <input type="checkbox"/> After collection of pre-assessment data (before unit) <input type="checkbox"/> After collection of post assessment data (after unit) <input type="checkbox"/> After unit
<b>STEP 6 - Reporting results</b> <input type="checkbox"/> Sharing of action research with cohort group	<b>Timeline:</b> After analysis of data and reflection
<b>STEP 7 – Taking informed action</b> <input type="checkbox"/> With input from your cohort group, determine needs on which to take action – determining your plan of action will inevitably lead to starting the action research process again.	<b>Timeline:</b> After reporting results

Of course, the primary purpose of the data is to guide the EPP in evaluation of its preparation efforts. It is useful for CAEP purposes only if it is useful to the EPP.

Assessment of the case study for purposes of CAEP review will include four distinct rubrics that follow from the Teacher Work Sample completed during student teaching. These include:

1. Design for Instruction
2. Instructional Decision Making
3. Analysis of Student Learning
4. Self-Evaluation and Reflection

EPP-constructed evaluation rubrics for the teacher action research projects. This is one on analysis of student learning.

Excerpt from the evaluation rubrics:

**3. ANALYSIS OF STUDENT LEARNING** - *The teacher uses assessment data to profile student learning and communicate information about student progress and achievement. The assessment data must be organized around your original learning goals.*

#### **Task**

Analyze your assessment data, including pre/post assessments and formative assessments to determine students' progress related to the unit learning goals. Use visual representations and narrative to communicate the performance of the whole class, subgroups, and two individual students. Conclusions drawn from this analysis should be provided in the "Reflection and Self-Evaluation" section.

#### **Prompt**

In this section, you will analyze data to explain progress and achievement toward learning goals demonstrated by your whole class, subgroups of students, and individual students.

- **Each learning goal should include an analysis:**
  - State the original goal;
  - State to what degree the students achieved it;
  - Provide multiple forms of evidence to support the achievement claims; and
  - Provide an explanation of how students performed in relationship to the original assessment criteria as described in the assessment plan.
- **Whole class.** To analyze the progress of your whole class, create a table that shows pre- and post-assessment data on every student on every learning goal. Then, create a graphic summary that shows the extent to which your students made progress (from pre- to post-) toward the learning criterion that you identified for each learning goal (identified in your Assessment Plan section). Summarize what the graph tells you about your students' learning in this unit (i.e., the number of students met the criterion).
- **Subgroups.** Select a group characteristic (e.g., performance level, socio-economic status, language proficiency) to analyze in terms of one learning goal. Provide a rationale for your selection of this characteristic to form subgroups (e.g., high- v. middle- v. low-performers). Create a graphic representation that compares pre- and post-assessment results for the subgroups on this learning goal. Summarize what these data show about student learning.
- **Individual.** Select two students that demonstrated different levels of performance (e.g. socio-economic status, language proficiency). Explain why it is important to understand the learning of these particular students. Use pre-, formative, and post-assessment data with examples of the students' work to draw conclusions

about the extent to which these students attained the two learning goals. Graphic representations are not necessary for this subsection.

### **EXAMPLE #3: PUBLIC University—P-12 student value-added data as part of state teacher evaluation, complemented by planned teacher action research**

Public, IHE enrollment around 20,000; EPP enrollment around 3,000

PUBLIC University demonstrates the impact of its completers on P-12 student learning and development, classroom instruction, and schools... It documents, using multiple measures that program completers contribute to an expected level of student-learning growth. Multiple measures include growth measures (e.g., value-added measures) required by the state for its teachers and available to PUBLIC's EPP, and other state-supported P-12 impact measures...

VAM measures reported by state

- Evidence for this standard includes PUBLIC's value-added data reported by the state board of education.
- Value-adding Teachers from the "Higher performing" and "Above average performing" categories in our state who have teaching licenses recommended by PUBLIC University comprised 23% of those employed as teachers in the years 2010 to 2013, 26% in 2013, and 37% in 2014.
- In the years 2010 to 2013 those who achieved an "Average performing" value-added classification comprised 45% of the PUBLIC graduates teaching in our state, 35% in 2013, and 44% in 2014.
- The percentage of PUBLIC graduates in the "Approaching average performing" category in 2010 to 2013 was 15%, 18% in 2013, and 18% in 2014.
- Those graduates in the "Least effective" category comprised 17% in 2010-2013, 20% in 2013, and 11% in 2014.

EPP includes self-study report information to describe the qualities of the state data and the results.

These are the EPP's conclusions as to what it can learn from the limited amount of state data on VAM. Note, however, that it is difficult to interpret results without more information on the representativeness of the samples and the differing sample sizes.

The data demonstrate higher value-added scores for the graduates from 2014 and a trend toward higher value-adding graduates over time.

Value-Added Data for PUBLIC University-Prepared Teachers							
Year	Teachers w/ Effective Licensure for the past 4 years		Associated Value-Added Classifications				
	Employed as Teachers	Teachers w/ Value-Added Data	Most effective performers	Above Average performers	Average performers	Approaching average	Least Effective
2010-2013	11367	2518	N = 316 13%	N = 264 10%	N = 1130 45%	N = 381 15%	N = 427 17%
2013	350	86	19%	7%	35%	18%	20%
2014	659	171	16%	11%	44%	18%	11%

PUBLIC demonstrates through structured and validated observation instruments, that completers effectively apply the professional knowledge, skills, and dispositions that the preparation experiences were designed to achieve.

The EPP could make interpretations clearer by including more information on comparisons with other teachers and more explicit descriptions of the VAM score classifications.

- Evidence for this element has been gathered from the results of the state teacher evaluation assessment. Our state’s system for evaluating teachers provides educators with a rich and detailed view of their performance, with a focus on specific strengths and opportunities for improvement. The system is research-based and designed to be transparent, fair, and adaptable to the specific contexts of the state’s districts. Furthermore, it builds on what educators know about the importance of ongoing assessment and feedback as a powerful vehicle to support improved practice. Teacher performance and student academic growth are the two key components of the state’s

The state teacher evaluation assessment is the source of the VAM data and also of other information that helps to put the VAM data in context so that the EPP can interpret those data.

Description of state system (paragraph above) followed by EPP description of limitations of the state data.

evaluation system. Limitations of these data include:

- The information in the report is for those individuals receiving their licenses with effective years of 2010, 2011, 2012, and 2013;
- The teacher evaluation data in this report are provided by the state education agency based on the original framework of 50 percent teacher evaluation and 50 percent student growth measure, and the number of teachers (N) with associated state evaluation assessment data remains small at this point, and due to state regulations, must be masked for institutions with fewer than 10 linked teachers.

Many states would have a mix of student-based data and other factors included together in teacher evaluations.

The data, which the EPP reports, for PUBLIC University graduates who have state teacher evaluation scores demonstrates that the percentage of skilled and

accomplished teacher evaluation respondents for 2010 was 93%, for 2011 was 100%, for 2012 was 82% and for 2013 was 85%.

This was an observation made by the CAEP site team.

NOTE: The formative feedback made the following point: The EPP provides some evidence of completer impact on P-12 learning, but these data are limited and appear not to include all program areas. PUBLIC University responded as follows:

To address the limitations of the state value-added data, we propose the following 4 options.

1. The first option is to accept that the data from the state value-added model as presented adequately meets the intent of this CAEP indicator. The data at present are all that the state education agency provides to institutions of higher education (IHEs). These data represent evidence from the sophisticated SAS®/Education Value-Added Assessment System (EVAAS) model used statewide for the past 12 years at the district/school level and for the last 5 years at the teacher level. It is an integral part of the state teacher evaluation assessment.

Here is an elaborated response from PUBLIC university in the form of 4 options they might consider to address limitations of state value-added data, together with their reasoning. While this is specifically a response to formative feedback, it is really a good example of what an EPP might think through as it conducts its self-study, prior to writing its initial submission.

The advantages of accepting as sufficient the value-added assessment data our state collects are:

- a. the data are readily accessible; they provided annually to the university free of charge per the state regulations;
- b. these data are highly reliable (Not all Value-Added Models are Created Equal; Battelle Kids; Eric Hanuschek; ODE);
- c. the EVAAS model is gold standard for decision making;
- d. every reading and math teacher in Ohio in grades 4-8 receives a value-added estimate annually so as such, our grads who are assigned in these grades and subject areas receive a value-added estimate.

The limitation of relying on these data to determine the effects of our programs are:

- a. these data represent only teachers in grades 4-8 in reading and math in our state; therefore, they are not representative of our programs, e.g., adolescent young adult, early childhood, or some special education; and cannot be generalized;
- b. a small percentage of PUBLIC University graduates who are novice teachers are represented in the sample.

2. The second option relates to expanding the sample to include extended value-added estimates. Our state allows districts the option to augment the state value-added model. Districts may opt to administer standardized, norm-referenced assessments that can be used to produce an alternate value-added model. This approach can produce value-added reports at the early elementary and high school levels by subject area. In addition, these reports may be generated for science and social studies in grades 4-8.

The advantages for using an alternative value-added model are:

- a. capturing these extended value-added results has the potential to provide us with information that better represents our programs, and
- b. would potentially produce a better representation of our novice teachers.

The limitations of using the extended estimates are:

- a. districts are not required to administer the assessments required by these analyses; consequently, there is no known match between the alternative value-added reports and our novice teachers;
  - b. we do not have permission, at present, to obtain these data from the Ohio Department of Education; therefore, we would have to enter into an agreement with the Department to obtain them;
  - c. there is a cost to generating an analysis of these data that would provide evidence of our graduates' effect on student growth.
3. Our third option is to use student learning objectives (SLOs) as data proxy for value-added estimates. The state education agency requires the teachers who are not eligible to receive value-added reports must generate student learning objectives. These SLOs are purportedly a measure of student growth created by teachers. We propose analyzing SLOs and resulting data with a partner district. The advantages of this option are:
- a. these data are potentially available to us contingent upon an agreement with the state education agency since these SLOs are required and collected by agency as part of the state teacher evaluation assessment; and
  - b. we would have a larger sample that would more closely represent our programs and our novice teachers.

The limitations of this option are:

- a. the use of these is contingent upon an agreement with the state education agency; and
  - b. there is no evidence to support the reliability and validity of SLO scores.
4. We recognize that value-added measures will not be available to most teachers in the foreseeable future because it would require testing all students in all years, in all grades, in all subjects. Current climate around over testing is a strong indicator that additional state testing that is needed to produce supplementary value-added estimates is

Ideally, the reasoning and this conclusion would have been in the initial submission, rather than as a result of formative review. While the EPP concluded that a single option was best in this situation, another EPP might conclude that a mix of the options would provide the most complete picture and balance strengths and limitations of individual measures.

improbable. Reliance on local district decisions around alternate value-added models is not likely to be sustainable and the weakness inherent in the use of SLOs creates a reluctance to utilize them in our accreditation process.

As such, PUBLIC University educator preparation programs are interested in tackling in a more meaningful way this challenge of measuring student growth. Our last option; therefore, would be to view the weakness in current SLOs as an opportunity to work with a partner school to

This is the EPP's preferred option

produce reliable, defensible measures of student growth using classroom data. The advantages of this option are:

- a. school faculty will utilize and university faculty will support sound assessment design principles and measures of student growth in P-12 classrooms;
- b. PUBLIC University can lead in the design of classroom-based student growth measures that produce meaningful diagnostic data to evaluate the effects of educator preparation programs; and
- c. the resulting data on student growth measures will be used to determine the effects of novice teachers on student learning. Not only will PUBLIC candidates and graduates benefit, but P-12 students and their families will benefit also.

We propose to continue to collect and report using option #1 while simultaneously focusing on proposed option #4. Together these options will produce a balanced approach to using value-added data, increase assessment literacy in classrooms, and result in a robust accountability system that benefits all stakeholders including university faculty and candidates, school personnel and most importantly, P-12 students.