

Report of the Data Task Force to the CAEP Commission on Standards and Performance Reporting

Peter Ewell

June 12, 2013

CAEP created the Data Task Force¹ to provide technical advice to the Commission on Standards and Performance Reporting about measures proposed to be used in conjunction with the five new Accreditation Standards that the Commission is recommending to the Board. Fifty-nine measures were proposed, grouped under the five Standards and as a free-standing group of eight measures to be reported annually to CAEP by accredited Educator Preparation Programs (EPPs). These measures were accorded equal prominence to the Standards themselves by the Commission, so issues such as the priority among measures, the details of their calculation and collection, and how these measures should be used in helping to make accreditation decisions were held to be extremely important. These issues were reflected in the charge to the Data Task Force².

The CAEP Data Task Force met in Washington D.C. on May 20-21 and came to a number of conclusions about the measures as reflected in this report. Two members of the Task Force were unable to attend the face-to-face meeting and their contributions were provided to the Chair by means of subsequent telephone calls. The report consists of five sections: a) establishing priorities among measures, b) general recommendations, c) recommendations on the eight Annual Report Measures, d) recommendations on selected other priority measures and, e) recommendations on using measures to inform accreditation decisions.

Establishing Priorities Among Measures. In order to structure the discussions at the meeting, the Chair of the Data Task Force and Study Director (Peter Ewell) and the CAEP staff member responsible for developing the Standards and associated review processes (Emerson Elliott) assigned a selected set of measures to an initial top priority group of twenty-three measures. These included eight “annual report” measures, plus a selection of others across the standards that represented critical factors in the standards and that, collectively, reflected a range of types (e.g., tests, impact on student learning, statistics, case studies, and curriculum). These assignments were then ratified by the Task Force. The Task Force believes that the highest priority should be accorded to student impact measures and measures of ability to teach effectively from the clinical point forward. They also agree that these priorities would help EPPs to determine what CAEP holds to be important in determining program quality and, hence, what EPPs should concentrate on. To structure the discussion on May 20-21, a further reduced set of fifteen measures were selected. These consisted of a) the eight Annual Report Measures and, b) seven other measures selected to address typical issues for different kinds of measures listed under the Standards.

¹ See Appendix A for a list of members.

² See Appendix B for the Charge to the Data Task Force.

General Recommendations. Based on its deliberations, the Task Force offers to the Commission a number of prefatory recommendations that cut across all measures. These include the following:

- In cases where efficiencies might be gained, EPPs should obtain measures from state education, higher education, or workforce agencies and CAEP should work with states to help establish the requisite data flows. A good example is in the area of employment rates, where state agencies have access to Unemployment Insurance (UI) wage record databases.
- In conjunction with the above, CAEP should describe and communicate to stakeholders each state’s data capacity and infrastructure including the volume and properties of its available data on employment, licensure and value added. This may stimulate those states that are not currently providing needed information to do so.
- Wherever possible comparative or standard measures should be used. Where this is not currently possible, CAEP should be steadily moving in this direction.
- While continuous improvement is a worthy goal of accreditation, its “bottom line” remains applying rigorous standards as part of an accreditation decision.
- Measures should be chosen and applied such that they conform as fully as possible to widely accepted characteristics of sound evidence (see the accompanying paper “Principles for Measures Used in the CAEP Accreditation Process”).
- All measures need triangulation through the use of multiple sources and methods.
- The eight Annual Report measures should be prominently displayed on EPP websites and also reported by CAEP.

Recommendations on the Eight Annual Report Measures. The Task Force believes that all measures used for CAEP accreditation require consistent definition, faithful representation, regular recurrence, and common reporting protocols. That said, the Task Force made recommendations on each of the eight Annual Report Measures as follows:

Completion Rates. The Department of Education has established a standard method for reporting institutional graduation rates. This method is embodied in the Graduation Rate Survey (GRS), which is a part of IPEDS. The calculation is based on a cohort of first-time full-time students which is tracked through a period corresponding to 150% of the nominal time required to complete the degree (e.g. six years for a four-year program or 3 years for a two-year program). This rate is already familiar to higher education institutions so it is logical to use it wherever it applies. Unfortunately, there are a couple of drawbacks to doing so:

- Late Entry. As undergraduate majors, many teacher preparation programs are structured so that students enter them after initial enrollment at the institution—most commonly at the end

of the sophomore year. If they all enter the program at the same time and are allowed two years to finish, a three year tracking window can be created. But if they enter at different points in time this matter can become complicated. This is also the case for students transferring into teacher preparation programs from other, partially completed, majors or from other institutions.

- Graduate Programs. Some teacher preparation programs are offered at the Masters level so students enter them after completing an undergraduate degree. The good news is that this provides a single point of entry upon which to base a cohort for the teacher preparation program. The less good news is that there is no national standard against which to judge the adequacy of such a rate, although the 150% of catalogue length criterion would be a good place to start.
- Part-time Attendance. Students may also enroll on a part-time basis so will naturally take longer to complete. At present, there are no established standards about how long to allow such students to graduate. Faced with this situation, institutions with large numbers of part-time students frequently adjust the amount of time allowed proportionate to the load students carry. The Voluntary Framework of Accountability (VFA) established by the American Association of Community Colleges (AACC) may provide a good model for moving forward here.
- Certification Students. Many teacher preparation programs provide continuing professional development to already-employed teachers through a range of short certification programs. These “students” are already teaching, so are not seeking a degree. Despite their presence as enrollees, the Task Force recommends that these students not be included in any graduation rate calculations.

Given these complicating factors, the Task Force believes that several appropriate graduation rates can be calculated including various cohort methods, supplemented by ratio methods (graduates/number of students enrolled). In addition, the Task Force recommends that a standard rate based on the 150% of program length criterion be applied to junior-entry and Masters-level programs, recognizing that there may only be a few of these. It is an eventual goal for CAEP to evolve a standard methodology for all EPPs to report on, but this remains at the moment an aspiration. Graduation reporting in the near term may require EPPs to provide their own measures, describe how they are calculated, and frame an argument about why these alternative measures are sound and appropriate. After CAEP evolves a standard measure, EPPs can supplement this measure with one of their own, but only after providing a sound argument that theirs is valid and superior. The Task Force sees completion rates largely as consumer information and less an indicator of program quality and, as such, believes that these rates should be disaggregated by gender and race/ethnicity. EPPs should also work toward improving these rates over time.

Licensure Pass Rates. The Task force recommends that these be provided for all fields that the EPP’s state requires and that they should be provided for multiple years. The Task Force recognizes that at

the moment not all licensure scores or pass rates are comparable, but CAEP should work with states and CCSSOs toward the eventual goal of comparability. This can happen now across states using Praxis tests, but not for state-specific Pearson tests, which will only allow comparisons within states. The new on-line Pearson test is common across states so if EPPs are subject to it, multi-state benchmarking is possible. The Task Force recommends that CAEP apply a single cut score for calculating success rates despite the fact that cut scores currently differ among states. The Task Force further recommends that the Commission ask CAEP to form a working group to work toward comparability and benchmarking. Finally, the Task Force recognizes that current licensure scores do not for the most part predict later student performance as teachers, so they should not be used as an indicator of program effectiveness on their own.

Employment Outcomes. The Task Force recognizes that the employment status of completers is not ascertained or reported in many cases but recommends an aspirational goal of doing this on a comparable basis across EPPs. Reporting categories should include a) employed anywhere, b) employed in education, c) employed in field of preparation, d) pursuing further education and, e) none of these. CAEP should establish regional benchmarks for each of these status categories to reflect local labor market conditions. Status should be reported for completers after one year, two years, and five years. The Task Force recommends that CAEP work primarily with state education and workforce agencies to obtain the necessary data about employment. While completer surveys are the dominant current mode for collecting information on employment outcomes, states should move toward harnessing UI-Wage records to track employment and the National Student Clearinghouse to track graduate school attendance.

Loan Default Rates. The Task Force does not believe loan default rates are a valid measure of program quality and does not support their use in making accreditation decisions. It does see value in this measure as part of a larger package of consumer information, together with such things as first-year earnings and program costs.

Surveys. The Task Force discussed three kinds of surveys—of completers, of employers, and of pupils of graduates. The first two should be mandatory for all EPPs and the third should be optional. In collaboration with states, CAEP should work to develop common surveys using those in such states as Ohio and Missouri, or those developed by TEAC, as models. Response rates on all surveys should be reported, although no minimum response rate should be specified. Surveys should be conducted regularly (at least bi-annually) and should involve random samples of sufficient size (although the universe would be preferred). Common benchmarks for comparison should be established by CAEP.

Classroom Observation. The Task Force recommends the use of classroom observation measures but emphasizes that their results depend upon rigorous training on properly designed protocols. EPPs should consequently report in detail how the observations were conducted and how training was provided to classroom observers. Because the validity of observational data depends so heavily on training and appropriate protocols, the Task Force believes that they require triangulation with the results of other data collection approaches. The Task Force recommends that observational protocols

address both teacher and student behaviors. Technology-enhanced approaches such as the use of avatars as “simulated pupils” should also be explored.

Value-Added Measures. The Task Force supports the proposed change in terminology to refer to these as “impact on P-12 learning.” The Task Force recommends use of these measures as a conceptual advance over more traditional measures, given the aspiration to disentangle the various contributors to teacher effectiveness. The Task Force points out that use of these measures in the aggregate to make assessments of program quality is more valid and reliable than their current use in judging individual teacher performance. While a minimum number of candidates could be specified, the Task Force believes that this standard should be set empirically in each instance based on the observed ratio of signal to noise in the data. Most of the available empirical work examining these measures has focused on English and mathematics, but the Task Force supports expanding the work to other subject areas. CAEP should also make EPPs and teams aware of the NRC report on Value Added Measures and the caveats it provides about using such data. Finally, the Task Force cautions CAEP to use only true value-added measures; not all measures called “value added” employ genuinely longitudinal or pre-post methods based on equivalent examinations.

The Task Force is also aware of the fact that not all states collect such measures and no state collects them in all areas or grades. In the absence of such measures, the Task Force recommends that CAEP insist that EPPs provide direct evidence about candidate impact on student learning. To reach this goal, the Task Force recommends that EPPs gather such evidence from pre-service settings, if these data are not already available from in-service sources such as state teacher evaluations. With regard to the latter, the Task Force notes that alternative preparation providers now predominantly have arrangements with districts such that candidates serve as “teachers of record” for the period of their preparation. For these students state data on the learning associated with a particular teacher should be available. The Task Force recommends that all of this be supported by a credible process of collecting and adjusting measures, including adjustment for teaching experience, where appropriate.

Recommendations on Selected Other Priority Measures. The Task Force discussed the other seven priority measures selected for review and made the following recommendations and observations:

1.c. Capstone Measures. The Task Force notes that there are currently no common capstone measures in place but a number of them are promising. The most fully developed is EdTPA which was built under the auspices of Stanford University and the American Association of Colleges of Teacher Education (AACTE) and is now offered in partnership with Pearson. Assessments cover most domains of teacher preparation. Twenty-four states and the District of Columbia are participating in EdTPA, four of which have a state policy in place. The assessment will be available for use in the 2013-2014 academic year. The Task Force believes that another promising measure is the ETS Praxis Pre-Service Portfolio, which will launch in the fall of 2014. It is aligned to the Common Core, InTASC, and state standards and can be embedded in teaching and clinical experiences. The portfolio comprises four tasks—knowledge of students and the learning environment, assessment and data collection to measure student learning, instruction and technology, and a culminating task video. Until such common capstone measures are available, programs will continue to report results of their own capstone measures. For most of them

this will be based on a portfolio containing exhibits that demonstrate mastery of content or pedagogical knowledge. These portfolios are frequently aligned with the Charlotte Danielson framework for professional teaching practice, but many are also aligned with INTASC or state standards.

In the light of these developments, the Task Force believes that the desired objective is for CAEP to work with states toward the establishment of a set of common capstone measures based on one of these promising examinations (or any other promising alternative that may emerge) for all EPPs by 2020. All of these also require EPPs to provide evidence that they have validated the measures they use, and to provide information about how they train observers to use such examinations or frameworks reliably.

1.d. Impact on Student Performance in Clinical Practice. The Task Force notes that this measure is particularly relevant for non-traditional candidates. Consistent with its comments about Value Added Measures, the Task Force also observes that there may be multiple opportunities for student performance to be observed in the course of clinical placement.

1.g. Assessment Proficiency. The Task Force conceives of this as an input measure in which EPPs provide evidence that candidates have been exposed to coursework about assessment, have experienced assessment in real world settings, or provide information about the assessments used in their programs more generally. The Task Force notes that there is no current assessment that addresses assessment proficiency, but the ETS Praxis Pre-Service Portfolio has a task focused explicitly on this topic. Consequently, the Task Force recommends the adoption of the assessment proficiency measure as written. The only amendment the Task Force offers is to ensure that students are exposed to content about both formative and summative assessment.

2.g. Clinical Experiences. The Task Force finds this measure confusing in that the terms “high leverage practices” and “diverse settings” are undefined. The Task Force approves of the references to Ball, Forzani, and Danielson and recommends that these frameworks be used as the basis for framing evidence around this measure. The main focus should be on the extent to which clinical experiences demonstrate that candidates can teach effectively and can show positive impacts on student learning during clinical practice. With respect to “diverse settings,” the Task Force recommends that this term be better defined by the Commission to clarify which kinds of diversity are referenced. Whatever the eventual definition, the Task Force recommends that this measure be framed in terms of evidence that candidates are able to address whatever diversity presents itself in clinical experiences and that candidates have knowledge of the general principles involved in recognizing and dealing with any form of diversity.

3.e. Evidence on Candidate Quality During Preparation. The Task Force found the substance of this measure to be adequate for the moment. But it recommends that CAEP take steps to gradually move away from program-specific measures toward common measures like EdTPA, the ETS Pre-Service Portfolio, or the Renaissance TWS project.

5.f-g. Leadership Commitment and Stakeholder Involvement. The Task Force feels that the leadership commitment sections of this statement lack clear definition and recommends that they be clarified. Because other accreditors also have such standards, the Task Force recommends that CAEP look for

examples of rubrics or other ways these accreditors judge the extent to which leadership commitment is present.³ The Task Force recommends that stakeholder involvement be examined through inspection of membership lists on program advisory committees, together with minutes and agendas of committee meetings.

Recommendations on Using Measures to Inform Accreditation Decisions. The Task Force believes that it is important to emphasize that accreditation decisions are ultimately based on peer judgment. Decisions about accredited status should thus be based on examination of the *entire body* of assembled evidence about the condition and performance of a given program. The Task Force believes that this judgment should not be reduced to a mechanical process of calculation and ranking. The calculation or aggregation procedures that the Task Force suggests below should thus always be used to inform, not dictate, an accreditation decision.

In the light of this overall guidance, the Task Force believes that CAEP will need to address a number of issues in combining measures to help arrive at an accreditation decision for a particular EPP:

- The first is a determination of whether all measures are accorded equal priority. If certain measures are to count for more, they must be weighted accordingly.
- The second issue is how to aggregate measures to arrive at an overall decision. As noted, this is ultimately a holistic judgment, but several calculations may help guide this judgment.
 - One is to average the measures after standardizing their individual values. This method is compensatory, in that high performance on one measure can make up for low performance on another.
 - The alternative is to average only the measures within a given cluster of measures (for example, all of the measures included under a given Standard). This alternative is more rigorous with respect to quality assurance because it guarantees that the program meets all five standards. The Task Force therefore recommends this latter aggregation approach.
- A third issue is to determine a particular level within a given Standard above which the Standard will be considered to be met. “Cut scores” of this kind may be determined in many ways, but standard practice is to use a process of structured peer judgment (such as a modified Angoff or similar procedure) to make the determination.
- Finally, for measures that are collected at multiple points in time, triggers can be established if the value of any of them changes markedly. For example, a downward change of ten

³ Examples for the Western Association of Schools and Colleges (WASC) can be found at: <http://www.acswasc.org/googleresults.htm?q=program%20review>. Click on the entry “WASC Postsecondary Initial Visit Rubrics” and look for the rubric entitled “Criterion 2: Organizational Infrastructure and School Leadership.”

percentage points or more on the program completion rate for a given program might contribute to an accreditation decision or might trigger the program for an immediate full review.

After preliminary decisions are made on the above matters, the Task Force recommends that the method(s) chosen be thoroughly piloted for at least a year before formal adoption and should be tested and adjusted based on historical data, if they are available. Alternatively, it believes that CAEP could tentatively adopt two or more different approaches and pilot these on a sample of volunteer program.

Appendix A

List of Data Task Force Members

Peter Ewell, NCHEMS (Chair)

Wade Boykin, Howard University

Michael Feuer, The George Washington University

Kurt Geisinger, Boros Center for Testing (University of Nebraska)

Hap Hairston, Missouri Department of Education

Jillian Kinzie, Indiana University Bloomington

David Monk, The Pennsylvania State University

George Noell, Louisiana State University

Appendix B

Charge for the Data Task Force and its Relationship with the CAEP Commission on Standards and Performance Reporting

Draft; March 27, 2013, CAEP

The Council for the Accreditation of Educator Preparation (CAEP) needs an advisory group to assist in aligning accreditation evidence with standards so that rigor and innovation can be fostered by the new, unified, education accrediting body.

In its first large and publicly visible undertaking as the new education accreditor, CAEP created a Commission on Standards and Performance Reporting. That Commission is charged to create standards for a rigorous system of accreditation that demands excellence and produces educators who raise student achievement. The charge gives equal weight to “essential standards” and to “accompanying evidence” indicating that standards are met.

The Commission’s draft recommendations are currently available for public comment (see copy, attached) and, following analysis of the comments and one additional Commission meeting, will be completed by this coming summer. The Commission’s report includes examples of evidence for each of its recommended standards and includes recommendations, as well, for levels of accreditation that would be capped with an “exemplary” accreditation decision, signaling both a high quality of evidence that is valid for the claim being advanced, and that high performance markers are reached.

The Commission is asking for public comments on the appropriateness, rigor, comprehensiveness, and adequacy of its examples of evidence. In addition, the members believe that these examples need a close and more technical look that will assist their consideration of:

- “decision rules,” or the circumstances under which various kinds of measures can appropriately be used,
- thresholds to be set for type of evidence in each of the components of the Commission’s standards,
- guidelines for combining evidence about components and standards into a composite evidentiary base and applying that to assist judgments about different levels of accreditation, and
- guidelines that CAEP could follow to take account of differing data practices across states that influence the evidence available to preparation providers.

The Data Task Force is to advise CAEP on these data issues. This advisory role will broaden CAEP’s sources of counsel and provide additional expertise to assist the Commission as its members complete their task this year.

Peter Ewell is directing this effort for CAEP. Peter brings a wealth of experience to issues of higher education assessment and data, outcomes measures and performance-based certification, higher education indicators and measurement, and a long involvement with accreditation organizations, especially regional accreditors. A paper that we prepared for the Commission has been especially useful for the members as they have prepared their current drafts

The work of this advisory group, which we are labeling a “Data Task Force,” can use your time efficiently through exchange of papers and electronic meetings, preceded by an initial face to face meeting in Washington. The Task Force work needs to be accomplished prior to conclusion of the Commission, and May 31 is our target.