

Outcomes, Measures, and Data Systems:

A Paper Prepared for the CAEP Commission on Standards and Performance Reporting

Edward Crowe with Michael Allen and Charles Coble

Teacher Preparation Analytics, LLC

January 2013

The CAEP Commission on Standards and Performance Reporting aims to strengthen accreditation of teacher preparation programs in the United States by making preparation program data “more strongly performance oriented, consistent and complete for a sound evidentiary base.” The Commission’s charge includes “design of an annual report by which institutions can publicly demonstrate accountability and improvement.” Toward these ends, CAEP will develop standards to drive preparation program reporting, with a focus on outcomes such as candidate quality and content knowledge, as well as evidence from and about clinical experiences.

This paper was commissioned by CAEP to describe the data and data systems needed to support accreditation policies that foster outcomes-based teacher preparation. It discusses the extent to which current state education data systems have the capacity to collect, analyze, and report key indicators of quality for teacher education programs, their candidates, and their graduates. The authors also make a case for the outcomes that “outcomes-focused teacher preparation” should be all about, identifying measures for those outcomes as well as the current status of state-level data systems with respect to these outcomes indicators.

The CAEP Commission is reexamining teacher preparation accreditation at a time of tremendous change in American teacher education. Public confidence in teacher education is quite low, and new providers arrive on the scene almost every day. Efforts to improve teacher and program quality are pushed by foundations, the federal government, and the states. Existing quality control mechanisms are thought to protect weak programs and let too many underprepared teachers into the classroom. Furthermore, quality control in teacher education bears little resemblance to policies and practices in other forms of professional education held in higher esteem.

Since oversight of professional education and professional practice in the United States are primarily state-level functions—with some federal policy supports, incentives, and sanctions—CAEP also requested the authors to explore ways for accreditation and state policy to work more effectively to help states meet their obligations to K-12 students and taxpayers. We address this in two ways: with a brief contrast between teacher preparation and a few other professions, and by offering ideas through which CAEP itself can create incentives to improve teacher quality policy in the states.

Teacher Preparation Analytics, LLC is a consulting partnership focused on high leverage strategies to strengthen teacher and administrator preparation. TPA works with individual programs and institutions, state university systems, and with state and national organizations to develop action plans that address key problems and challenges. This work includes identifying measures, data, and methods needed for effective preparation program self-assessment and improvement; strategic support through research, convening, and capacity building for programs or networks; and research reports and policy briefs that analyze the impact of policies and practices on the quality and outcomes of educator preparation programs.

A Changing Landscape

Any serious effort to develop and use strong indicators of graduate and program performance must cope with several key issues: selection and measurement of key outcomes; identification of data sources; assessment of data quality and availability; and weaknesses in current practices that may inhibit realization of a successful outcomes-focused system of accreditation and preparation program improvement.

CAEP's goal—standards, accreditation policies, and program reporting practices built on data and measures of key outcomes—would be an important change from the current situation where accreditation and accountability does not guarantee program quality or meet the needs of K-12 schools and their students. Weak quality control for programs also undermines the reputation of teachers and the programs that prepare them for the classroom.

Current oversight and reporting mostly ignore the impact of program graduates on the K-12 students they teach, give little attention to where graduates teach, how long they remain in the profession, or the quality of classroom teaching. Nor is there any reason to think that teachers

who complete an accredited preparation program are more likely to demonstrate high-quality classroom teaching performance than those trained elsewhere.¹

The general landscape—few or no outcomes indicators for preparation programs across the country—is starting to change. The changes create real opportunities for CAEP, for state policymakers, and for teacher preparation programs interested in self-assessment and continuous improvement. The federal Race to the Top (RTT) program supports work in nineteen states and the District of Columbia “to encourage and reward states that are creating the conditions for educational innovation and reform.” With grants up to \$700 million, these states² are working on four strategies to improve teaching quality:

- Linking K-12 student achievement and growth data to the teachers of these students.
- Connecting this linked student-teacher information to all in-state teacher preparation programs.
- Developing public reports of program effectiveness for each preparation program in the state (all pathways into teaching).
- Expanding teacher education programs or teacher credentialing options that are successful at producing graduates who are effective teachers.

In pursuit of these goals, the states agreed to collect and report a wide range of outcomes-focused indicators for all of their in-state teacher education programs.³ Appendix B summarizes state commitments through Race to the Top.⁴ All this work and federal money mean that data collection and reporting capacity are being developed or enhanced in twenty states. In addition to developments supported through Race to the Top, states are taking other actions relevant to the mission and goals of CAEP: building capacity in data systems, reporting structures, software

¹ See Suzanne Wilson and Peter Youngs, “Research on Accountability Processes in Teacher Education.” In Marilyn Cochran-Smith and Kenneth Zeichner, eds., *Studying Teacher Education*. (Washington: AERA, 2005, p. 591-644).

² DC, whose Race to the Top grant is administered by the Office of the State Superintendent of Education (OSSE) is referred to as a state in this discussion for ease of reference.

³ The fact that large proportions of initially certified teachers in many states were prepared to teach in another state is a problem for this strategy, as it is for programs that send significant proportions (up to 70%) of graduates into other states. We discuss this challenge below in more detail.

⁴ This table and parts of the Race to the Top discussion are based on the analysis published in Edward Crowe, *Race to the Top and Teacher Preparation: Analyzing State Strategies for Ensuring Real Accountability and Fostering Program Innovation*. Washington, DC: Center for American Progress, 2011.

resources, and human capital expertise in ways that make it more likely that preparation programs in these same states will have a better chance to access and use quality information about their own outcomes, thus supporting their participation in the CAEP accreditation and reporting system. For instance:

- Some states are working on improved teacher licensure tests (measures of content knowledge, teaching knowledge, and teaching skills).
- Thirteen states have or will have preparation program satisfaction ratings for all teacher preparation programs from schools that hire their graduates.
- Eleven states are or will collect information about teacher evaluation ratings for preparation program graduates.
- Six states are said to be collecting data on pupil learning gains.
- Four states now collect and report retention rates in teaching of program graduates.⁵

What Do We Mean by Outcomes-Focused Teacher Preparation?

While the outcomes and measures described in this section are ideal ones that will take some time to realize fully, effective national accreditation and a solid teacher quality pipeline will be served best with a small set of outcomes that are measured transparently, reported openly, and employed for (and by) every program in every state. Given where teacher education is as a field, this may strike many readers as long-distant dream rather than a realistic scenario. Nonetheless, the authors believe concrete steps can and should be taken to make the ideal real in the shortest possible time. Without ambitious goals and a sense of urgency, deliberations and even the policies of organizations like CAEP may soon be irrelevant to the production and oversight of teachers in the United States.

In a sense, the outcomes-focused preparation programs envisioned by CAEP have to deal with two sets of outcomes: those related directly to teacher candidates and graduates as *individuals* (e.g., their knowledge, classroom performance, and persistence in the profession; academic outcomes for K-12 students taught by these graduates); and the aggregation of individual data into *program* outcomes. Some program outcomes simply sum up particular

⁵ For more on these state activities, see page 46 of the NCTQ State Teacher Policy Yearbook for 2011. This report can be accessed at <http://www.nctq.org/stpy11Home.do>. This analysis was completed and published before the state Race to the Top grants were fully operational, so some state RTT activities noted earlier may not be captured by the NCTQ findings.

measures of individual teacher performance, such as the proportion of program graduates who persist in teaching beyond one year, three years, and five years, the proportion of graduates who are effective teachers, or the number who pass content knowledge tests. Others are program-level indicators such as completion and employment rates, persistence in teaching, and satisfaction survey feedback from employers and graduates.

To the extent that university-based teacher education programs are outcomes-focused today, they are often constrained by what the state expects them to report. This is also true for non-university based programs. From the standpoint of outcomes-focused programs, therefore, CAEP's challenge is multi-faceted. Among the tasks that CAEP is likely to be grappling with are these three:

- Adopt outcomes as core drivers of accreditation and hope they become central to the operation of teacher preparation programs that seek CAEP accreditation.
- Construct procedures for CAEP review of outcomes performance as well as reviews of the curricular activities behind them for institutions that may well be grafting CAEP's outcomes onto what they already do. Given the "compliance mentality" that many in the field have adopted as a response to federal, state, or accreditor reporting requirements,⁶ the risk for CAEP is that programs will simply collect and report data on new indicators without reexamining core features of their programs. Thus, the only real program-level use for new indicators will be making sure the right stuff gets sent off to CAEP.
- Design and implement accreditation data collection and program review activities to capture post-hoc adoption of outcomes by programs with widely varying design features and operating characteristics (i.e., because programs vary so widely in their requirements and program features, it may be more challenging to map outcomes-focused accreditation policies onto preparation programs than it was to conduct more process-oriented reviews. Even so, the important thing is not how you get it done, but that you get it done.).

With regard to this last point, while it is perfectly natural for people to move things around from "old boxes" to "new boxes" (unless there is a new something for which they don't have a box), a

⁶ A commonplace observation of many preparation programs has been the extent to which they generate, archive, and report data because "they have to", with little consideration given to the quality of the data or what it tells the program about the ways in which candidates are prepared for teaching.

way to drive real change *within* programs, not just in their reporting, would be the use of carefully developed rubrics that describe the levels of compliance expected by CAEP. If there are some really new program behaviors expected by the revamped accreditation process, such as school-university partnerships, full year residencies, and the integrated use of master teachers, the Commission ought to construct and use strong rubrics that capture the quality of these preferred practices. This strategy can help turn compliance from a vice into a virtue, but only if program compliance is held to and measured against high standards.

Before accreditation policies are put in place to cope with these challenges, the first step is deciding which outcomes should drive the whole process. ***In other words, what outcomes should outcomes-based teacher education be all about?*** In discussing outcomes about teacher preparation programs for their national accreditor, measurement and data collection issues associated with each outcome are described in the body of the paper, as are some of the considerations important for data systems that can support outcomes-focused teacher education. A more extended description of these data and data system issues can be found in Appendix A.

Key “drivers” of teaching and learning ought to be at the heart of outcomes-focused teacher preparation. The choice of outcomes should communicate clear signals about the program’s goals and whether it does a good job in meeting those goals. CAEP can also select outcomes that communicate its own values and aspirations about teacher preparation as a form of high status professional education.

Some indicators are measures of *individual* candidates and teachers, such as program completion, persistence in teaching, or classroom observation ratings. On the other hand, there are measures that clearly relate to the *program* itself: survey satisfaction rates from all graduates and their employers; the proportion of program graduates who are employed in high demand subject areas or difficult to staff schools; or the value-added gains shown by program graduates who teach a certain subject or grade level. For the most part, however, it seems to the authors that many of the program-level indicators are summary measures of individual-level performance. If the point of outcomes-focused teacher preparation is to base accreditation decisions on program performance, measures of program performance are largely composed of the aggregated indicators of how individuals perform on tests, in the classroom, or in the profession. To take an example from another realm, hospital outcomes are measured at the

individual patient level (morbidity, mortality, infections, etc) and at the institutional level: **rates** of mortality, average days in the hospital, average number of readmissions, and so forth.

What the Wallace Foundation called performance measures that are “potent” behaviors linked as directly as possible to student learning means that empirical measures have to be the building blocks of outcomes-focused teacher preparation.⁷ Data collection, performance calculations, and reporting practices have to be transparent. Outcomes measures must meet standards of rigor and quality to inspire confidence. These are important traits of an outcomes-focused *system* so that education professionals on the inside will feel confident that all programs are being judged by the same rules. Policymakers and the informed public must also believe that accreditation and other preparation program decisions like state approval are based on outcomes data with important and accurate information. The confidence of other audiences also matters greatly: would-be teachers deciding where to enroll in a preparation program; successful teachers recommending their programs to the next generation of teacher candidates; K-12 schools seeking partnerships with teacher preparation programs; the public and the policy community respectful and supportive of teacher education; and public schools making teacher hiring decisions with predictably positive results for their students.

Classroom teaching skills

The classroom teaching performance of program graduates is a key outcome that programs, accreditors, and states ought to use as a quality measure. We describe this outcome measure ahead of student achievement and academic growth only because observation measures can be more easily collected from a larger number of teachers (or from representative samples of program graduates across the grades and subject areas). Data on student performance are less widely available because most teachers are assigned to untested subjects and grades. It may be useful (and cost effective) to explore ways to collect multiple measures of teacher performance from employers and mentor teachers on the performance of beginning teachers. That is to say, as more districts do better evaluation of their teachers, these school- or district-based data may be good sources of information for programs about graduate’s performance as teachers, if the district will share with the program their findings about graduates of the program who teach in

⁷ See the Wallace Foundation, “Assessing the Effectiveness of School Leaders: New Directions and New Processes” (2009), p. 5. Available at <http://tinyurl.com/ydv6ltc>.

the district. While the value of this information would be affected by the quality of district-based evaluation mechanisms, it might be worth looking into as a source of data.

Classroom observation and assessment of on-the-job teaching should be regarded as a key program outcome because no single measure tell us all we need to know about a program or its graduates. Some *programs* now employ classroom observation to gauge development of requisite knowledge and teaching skills by their teacher candidates, suggesting there might really be two performance-related measures here for outcomes-focused teacher education programs: performance of candidates *during* the program and their performance as teachers of record upon completion of the program.

Such data would help the program faculty and administrators identify knowledge and skill sets that make a difference in the professional practice of their candidates and graduates. Classroom assessment results can highlight areas for individual candidate improvement, and for preparation programs that provide induction support to new teachers, teaching assessment findings can flag areas where continued development of teaching skills would improve a graduate's overall effectiveness in the classroom and persistence in teaching.⁸ Widespread implementation of a classroom teaching performance outcome measure would be a major step in providing robust and relevant evidence about the connection between teacher preparation and student achievement.

It is important to bear in mind, however, that a system of quality classroom observation must support fair judgments based on reliable and valid findings for individual teachers and for groups of teachers.⁹ Not all classroom teaching observation protocols are the same. It appears as though few of those now used by teacher education programs (including most of those mandated by state regulations) meet standards of rigor. Candidates, graduates, programs, and the public deserve “validated, standardized observational assessments of teachers’ classroom instruction and interactions.”¹⁰

⁸ Gary T. Henry, Fortner, C. Kevin, and Bastian, Kevin C. (2012) “The Effects of Experience and Attrition for Novice High School Science and Mathematics Teachers” *Science* 335, 1118-1121.

⁹ Laura Goe, Courtney Bell, and Olivia Little. “Approaches to Evaluating Teacher Effectiveness: A Research Synthesis” (Washington: National Comprehensive Center for Teacher Quality, 2008), p. 9.

¹⁰ Robert Pianta and Bridget Hamre, “Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity” (*Educational Researcher* 38 (2), 2009), p.109.

Practical Issues

Fortunately, there are a growing number of quality classroom observation instruments available.¹¹ National studies and pilot projects are building a foundation of knowledge for using classroom observation as a program outcome. Two large studies have produced relevant findings by examining links between observation instruments and pupil learning gains through videotaped observations of many teachers.¹² Similarly, the edTPA initiative has pilots in 21 states with 7,000 teacher candidates from cooperating university preparation programs, with the focus on teaching skills while candidates are still in their programs.¹³ Advocates of edTPA hope it will be a reliable and valid source of performance information, but performance measures for program graduates are still needed.

Observational findings for individual program graduates have to be aggregated and summarized for all the graduates of a specific program in order to constitute a program outcome indicator. An alternative strategy would require large enough samples of graduates to produce reliable findings. Some programs do this on their own, using the evidence to guide candidate development and for program improvement. MET and other efforts can provide useful lessons, especially as states and districts implement higher quality teacher evaluation systems. Tapping these state and district datasets for program purposes (not to evaluate individual graduates) ought to be a focus of CAEP's work.

Teaching effectiveness

To many people, *the most important preparation program outcome is teacher effectiveness—the extent to which program graduates help their K-12 students to learn*. Since high quality instruction is the main in-school driver for student achievement, it makes sense that

¹¹ See the discussion of these issues can be found in Pianta and Hamre, “Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity,” p. 111; as well as Goe, Bell, and Little, “Approaches to Evaluating Teacher Effectiveness: A Research Synthesis,” p. 22.

¹² The Understanding Teacher Quality initiative examines six instruments through videotaped observations of 450 teachers, while the Measures of Effective Teaching (MET) project has videotapes for about 3700 teachers. For more information about the Understanding Teacher Quality project see <http://www.utqstudy.org/index.html>; more about the MET effort can be found at <http://www.metproject.org/index.php>.

¹³ See <http://edtpa.aacte.org/about-edtpa>.

teacher effectiveness measures ought to be the central outcome. Today, however, only a few states have elevated teacher effectiveness as a core expectation or outcome for preparation programs. Louisiana uses value-added analyses of student academic performance to make decisions about the quality of every public or private “traditional” or other pathway into teaching.¹⁴ A few years ago, Florida began measuring and ranking its teacher education programs according to the learning gains demonstrated by K-12 students taught by program graduates.¹⁵ And Texas has announced a program accountability policy that, like Florida and Louisiana, includes program graduate impact on K-12 learning as a core indicator.¹⁶ Tennessee and North Carolina have published studies linking prep programs to student achievement results but neither state uses the information for accountability or program improvement.¹⁷

Louisiana has had the longest track record as a state in using teaching effectiveness as a required preparation program outcome. It is still unclear how Florida and Texas will implement their policy focus on this outcome, and the work of the Race to the Top states (including Florida) with student achievement as a program outcome has not yet produced any publicly accessible reports of program performance.¹⁸

Practical Issues

States have relatively little experience with implementation of teacher effectiveness as a preparation program outcome, but at least 20 states have taken steps in this direction (19 Race to the Top states, including Louisiana in Round 3, plus Texas). Whether or not program faculty and administrators share this state goal, analyses and judgments will be made about programs in these states based on their performance on this indicator. This poses opportunities as well as challenges: improved state data systems are needed to link teacher and student data; effective

¹⁴ For more information on Louisiana’s system as well as the policies and research behind its development, see State of Louisiana, Board of Regents, “Teacher Education Initiatives,” available at <http://tinyurl.com/27y5fzg>.

¹⁵ More on Florida’s efforts can be found at <http://tinyurl.com/yjwd8md>.

¹⁶ Details on the Texas approach come from Senate Bill 174 and Chapter 229 of the Texas Administrative Code, both adopted in 2009. See <http://tinyurl.com/yz9jmfg>.

¹⁷ By this we mean that neither state’s education agency has yet found a way to incorporate the results of these analyses into their decisions about which programs should be authorized to remain open.

¹⁸ Tennessee began producing its annual performance reports and posting them on a website before receiving its Race to the Top grant. It’s not clear whether the state will actually do anything with this information.

confidentiality and privacy policies are crucial; and analysis of K-12 testing data must be careful to use appropriate statistical models.¹⁹

Many preparation program graduates in these states and across the country teach grades and subject areas that are not tested by the states; one estimate is that about two-thirds of teachers fall into this category. A major challenge, therefore, is to develop learning outcomes for students of teachers in these untested subjects and grades. CAEP and others interested in this problem can tap work underway by Race to the Top states that face the same problem and are trying to address it.

With respect to data systems needed to collect and analyze teacher effectiveness information, most states can link student and teacher data in their K-12 system, but they are not able to tie employed classroom teachers back to their in-state preparation programs. This will need to be worked out for accreditation and accountability, and it's also needed for programs themselves to acquire, use, and report information on the teacher effectiveness of their graduates.

Despite the challenges, value-added analyses and growth model calculations of student learning are becoming more common as states and districts work out ways of measuring student outcomes in order to improve them. Expanded use of these analytical strategies has stimulated efforts to improve the student tests that function as dependent variables, and it seems safe to say that the nation will see further work to refine the analytical methods used to determine the impact of teachers on the academic achievement of their pupils.

Candidate and teacher knowledge

Preparation program accreditation and accountability must take account of the content knowledge and professional knowledge of teacher candidates and program graduates. The problem is finding measures in both areas that are strong, credible, and useful indicators. Praxis and similar tests have been used by the states for many years, but few outside the profession see these tests—in their current incarnations (paper-and-pencil, non-performance based)—as credible indicators of candidate or new teacher knowledge. Many inside the profession share these doubts.

A recent report noted that more than 1100 teacher tests are in use across the fifty states, with over 800 content knowledge tests alone. Even when two or more states employ the same

¹⁹ Goe et al., “Approaches to Evaluating Teacher Effectiveness: A Research Synthesis.”

test of content or professional knowledge, the states set different passing scores.²⁰ According to the US Department of Education, 97% of all test-takers in the United States get passing scores on the current panoply of teacher tests.²¹

These are serious problems for quality control by accreditors and states, but the current teacher testing system also undermines public respect for teacher education and for the graduates of teacher preparation programs. This is because the current testing system for teacher candidates and program graduates has three significant flaws:

- There are too many tests being used to support a coherent structure of candidate/teacher knowledge assessment and quality control of programs as well as teachers.
- Passing scores for these tests are set too low to ensure that those who pass have the content and professional knowledge to be effective classroom teachers.
- The tests themselves have little demonstrable relationship to the knowledge, skills, and teaching performance required in today's schools.

To take the last point, the expert panel convened by the National Research Council reported in *Testing Teaching Candidates: The Role of Licensure Tests in Improving Teacher Quality* (Washington, DC: National Academy of Sciences, 2001):

“Teacher licensure tests focus on the knowledge and skills identified by panels of educators as critical for entry into the profession. They cover the material considered to be *minimally* necessary for beginning teaching. Teacher licensure tests are not designed to distinguish moderately qualified teachers from highly qualified teachers. They are not constructed to predict the degree of teaching success a beginning teacher will demonstrate” (NRC, p. 47).

Other researchers report “little evidence that...a relationship exists between teachers' scores on such tests and their teaching success.”²² Candidates who cannot pass these tests probably should not have been admitted to a program in the first place, and programs with low pass rates should be closed. But other than using teacher test data to set a much higher quality

²⁰ Crowe, “Measuring What Matters: A Stronger Accountability Model for Teacher Education,” p. 8.

²¹ “Preparing and Credentialing the Nation's Teachers: The Secretary's Eighth Report on Teacher Quality, Based on Data Provided for 2008, 2009, and 2010.” (Washington: U.S. Department of Education), p. 54.

²² Suzanne Wilson and Peter Youngs, “*Research on Accountability Processes in Teacher Education.*” In Marilyn Cochran-Smith and Kenneth Zeichner, eds., *Studying Teacher Education*. Washington, DC: AERA, 2005, p. 592.

floor than is currently the case in any state, the licensing tests now in use do not measure outcomes relevant to the academic success of K-12 students or their schools.

However, the real problem isn't testing. The fundamental issue for credible and effective quality control is the tests themselves and how they are used. Better tests—linked to vital teaching knowledge and K-12 learning outcomes—would make a significant contribution to understanding the outcomes of preparation programs. Moreover, if teacher education followed the example of some other professions a battery of high-quality tests of teacher knowledge, skills, and abilities could be adopted by every state using the same passing score criteria. This step would create a standard, easily grasped framework for program accountability. We already do this in fields like nursing, engineering, accountancy and medicine without infringing on state autonomy, or breaching the principle of federalism. We will say more about this later in the discussion of the relationship between accreditation and state policy.

Despite the issues associated with current efforts to measure candidate and teacher knowledge, there is a pressing need for better ways to capture this information. Making headway on this challenge would be a significant contribution to teaching quality in the United States *and* would help to enhance the professional status of teachers and the programs that produce them. A recent report from the Council of Chief State School Officers (CCSSO) shows that states may be ready for real reform in this regard. The Chiefs are calling for a multi-state effort to develop “innovative licensure assessments” that include evidence about teacher impact on student achievement. Their report also argues for state program approval standards that address a program’s ability to produce graduates who positively affect student learning.²³

CCSSO argues for a range of indicators to measure teaching and program quality—including observation data, pupil achievement measures, surveys of graduates and school leaders, program retention rates, and placement into hard to staff teaching positions. Perhaps of equal importance, the Chiefs call for state data on preparation programs, disaggregated in various ways, to be provided to accreditors.

The CAEP Commission has an opportunity to advance program quality, teacher quality, and the overall status of both by taking a stand on the tests, cut scores, and passing rates that will

²³ This discussion draws on “Our Responsibility, Our Promise: Chief State School Officers’ Task Force Report on Transforming Educator Preparation and Entry into the Profession.” (Washington: Council of Chief State School Officers, 2012).

be acceptable for programs to be accredited.²⁴ For example, CAEP should require uniform use of the same tests for every program seeking accreditation, no matter what state it is located in, accompanied by uniformly high passing cut scores applied nationally, no matter what an individual state might establish as its own passing score. CAEP is, after all, a national accrediting body. Its standards should set a high minimum expectation for programs seeking the distinction of being ‘nationally accredited.’ Currently, being nationally accredited is thought by some not to distinguish between programs otherwise seen as strong or weak. One of this paper’s authors has argued elsewhere that the number of teacher tests can be reduced by 90% in order to provide clear signals about program and graduate quality.²⁵

CAEP should draw from the experience of other professions when it comes to tests of content knowledge and professional knowledge. Engineering, accountancy, nursing, and medicine operate with uniform state accountability standards and requirements. In nursing, for instance, the NCLEX-RN is accepted by every state as the single licensure test that determines whether or not a program graduate is granted a license to practice nursing. Every state uses the same passing standard, and pass rates are tied to program accountability for more than 1200 professional nursing programs in the United States (<https://www.ncsbn.org/nclex.htm>).

There is a similar story in engineering. All states employ the same battery of tests for would-be engineers, and every state employs the same passing score (see <http://www.ncees.org/Exams.php>).

Medical licensure standards in the United States can be summarized in one chart at this link (http://www.fsmb.org/usmle_eliinitial.html) because there is agreement across the states and within the profession about the standards for entry into the profession and about standards of quality for medical preparation programs. The profession of accountancy follows a similar pattern, with all states using the same four-part Uniform CPA Examination and passing scores (see <http://www.bls.gov/oco/ocos001.htm#training>).

Nursing and the professional preparation of nurses have many similarities to teacher education, given its focus on clinical practice, so program oversight practices may be particularly

²⁴ A recent AFT report, “Raising the Bar: Aligning and Elevating Teacher Preparation and the Teaching Profession” (Washington: American Federation of Teachers, 2012) calls for a “universal assessment process for entry” but says nothing about passing standards or establishing a common passing score across all states.

²⁵ Crowe, “Measuring What Matters: A Stronger Accountability Model for Teacher Education,” p. 3.

relevant to consider. Like teaching, nursing is a predominantly female profession with multiple preparation pathways (e.g., hospitals, community colleges, universities) and more than one thousand different *providers* of nursing education. Nursing has a rapidly growing knowledge base. Nurses, doctors, and health care institutions now are grappling with the implications of “evidence-based medicine” for their practice and for the education of would-be practitioners. The academic quality of entrants into the nursing profession is not dissimilar from teaching.

For nursing and the other professions discussed here, the essentially national character of their standards and practices is a form of quality control that ought to exist in teacher education. The high degree of consensus within each field about the values, standards, and practices that define professional preparation is a mechanism that links accreditation, state oversight, and professional licensure. This strategy protects the public with the same set of rules in every state. It brings higher levels of public respect for the profession as a whole and for who serve the public through their professional work.

Practical Issues

States guard their roles in the control of public education, although they seem less troubled about accepting federal money for schooling than about following federal requirements for use of the funds. There is a distinction to keep in mind between federal standards, which we do not argue for here, and national professional standards that all programs ought to be judged by if they seek the sanction (and benefit) of national accreditation. Federal or national professional standards cannot trump state legal authority over the operation of K-12 education, but states can choose to adopt a uniform set of standards (e.g., accrediting requirements, licensure tests, test passing cut scores) *voluntarily* as they have in medicine, nursing, engineering, accountancy, and other fields.

In terms of access to data about test results, the current situation is that in one way or another, most states have information about the licensure test scores of candidates and teachers, collected by test area and at the individual level. Federal Title II report card requirements have led to improved data collection in this area. However, this information sheds little light on program outcomes because of test quality and testing practices. Until better tests and more robust state policies are in place, CAEP can move the dial in a significant way even within the current testing system. Regardless of whether the state

where a program is located takes these steps, CAEP can adopt these policies for accepting teacher tests as measures of content or professional knowledge:

- To be accredited nationally as a teacher preparation program, the program's graduates must pass relevant tests using uniform national passing cut scores set at the 75th percentile for all test takers in the nation. Setting a high bar at this level would ensure that only the strongest candidates would be allowed to enter the profession. An alternative to the 75th percentile is to set passing cut scores one standard deviation above the mean for all national test-takers. For a normal distribution of test score values, this would be about the 68th percentile.²⁶ In effect, the 68th percentile standard means that candidates would have to score in the top third of all test-takers.
- If a state established passing cut scores *higher* than the 75th (or 68th) percentile, CAEP should use either the state's score or CAEP's national cut score, *whichever is higher*.
- At least 80% of all program graduates would have to pass all relevant tests at this 75th percentile passing score, or at the cut score established through procedures discussed in the preceding bullet.
- Pass rate data and its calculation must be made transparent to CAEP and to the public, ending the practice of reporting pass rates only for "program completers".²⁷

Teacher retention and employment

Two outcomes related to the impact of preparation programs on K-12 schools are: how long graduates persist in teaching and where they are employed as teachers.²⁸ Studies and reports over the last decade have documented the impact of teacher turnover on schools and

²⁶ See <http://sociology.about.com/od/Statistics/a/Normal-Distribution.htm> for a discussion of a normal distribution and the proportion of cases falling within one or more standard deviations of the mean.

²⁷ To protect the brand name of the new CAEP organization and to secure its credibility within the policy community, the bar for passing teacher and candidate tests should be set to a level that CAEP thinks is valid. That standard ought to be applied nationwide. National accreditation is an option; if states embrace accreditation for program quality control, they ought to stand ready to do what is necessary to reach the standard for all programs authorized within their borders.

²⁸ Through their Race to the Top work, some states have added an indicator for the subject areas taught by program graduates, hoping to create incentives and pressure on programs to concentrate output in fields like special education, ESL, and STEM, while reducing chronic overproduction in a field like elementary education.

students.²⁹ As the Consortium on Chicago School Research noted in 2009, “High turnover rates produce a range of organizational problems for schools...thwart efforts to develop a professional learning community among teachers and make it difficult to develop sustained partnerships with the local community.”³⁰

It has been widely reported that teacher turnover is a serious problem in low-achieving schools that have high proportions of poor and minority students. Teacher effectiveness studies show, however, that positive teacher impact on student achievement grows as teachers gain experience (up to a point), which mean that teacher turnover thwarts student academic performance. Research also indicates that preparation matters when it comes to teacher effectiveness.³¹ It is particularly important where candidates obtain their clinical experience during preparation, and it matters how a program’s clinical component is organized and supported by faculty so that graduates become effective teachers.³²

And yet high rates of teacher turnover persist despite the claims of many teacher preparation programs that their graduates are specifically prepared for challenging schools.³³ K-12 schools are already held accountable for the consequences of teacher turnover: high rates of turnover lead to weaker student academic gains than would otherwise occur. Preparation programs are not solely responsible for turnover or for its solution, but given the causes and consequences of

²⁹ Studies and reports on teacher turnover include work by NCTAF in *No Dream Denied* and their 2007 study of teacher turnover in five school districts (see <http://tinyurl.com/22wasx>), work by Smith and Ingersoll (2004), and the study of turnover in Illinois by White et al. (2008). More recently, the Consortium on Chicago School Research provided a very detailed analysis of teacher turnover and its impact of particular schools and students. See Elaine Allensworth, Stephen Ponisciak, and Christopher Mazzeo, “The Schools Teachers Leave: Teacher Mobility in Chicago Public Schools” (Chicago: Consortium on Chicago School Research, University of Chicago, 2009).

³⁰ Allensworth et al., “The Schools Teachers Leave.”

³¹ Boyd et al., “Teacher Preparation and Student Achievement”; Harris and Sass, “Teacher Training, Teacher Quality and Student Achievement;” the essays in Dan Goldhaber and Jane Hannaway, eds., “Creating a New Teaching Profession”. (Washington, DC: The Urban Institute Press, 2010); and

³² Pam Grossman et al., “Constructing Coherence: Structural Predictors of Perceptions of Coherence in NYC Teacher Education Programs.” *Journal of Teacher Education* (2008) 59; Donald Boyd, Pamela Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff, “Teacher Preparation and Student Achievement.” *Educational Evaluation and Policy Analysis*, (2009) 31 (4); and Matthew Ronfeldt, “Where Should Student Teachers Learn to Teach? Effects of Field Placement School Characteristics on Teacher Retention and Effectiveness.” *Education Evaluation and Policy Analysis*, March 2012, 34 (1).

³³ Many programs don’t know very much about whether their graduates become teachers or how long they stay in the profession. And few know whether their graduates teach in the kinds of schools the program believes it has trained them for.

teacher turnover, persistence in teaching is a program outcome that can help to align the interests of producers and employers.³⁴

Why should persistence rates matter as a program outcome? How can preparation programs and CAEP address teacher persistence rates? At least five states are working through Race to the Top on teacher persistence as a preparation program indicator (see Appendix A). CAEP has argued strongly for the “clinical residency” model of teacher preparation, for programs “that are fully grounded in clinical practice and interwoven with academic content and professional courses.”³⁵ Programs that take (or have taken) significant steps to implement a well-designed clinical residency model are likely to produce graduates whose experiences in a really rigorous clinical approach to preparation will provide them with the knowledge, skills, and teaching experience to survive school environments that are less than ideal. Better teacher preparation along these lines plus improved school working conditions are probably the keys to teacher retention.

Given the pervasive problem of teacher turnover, particular in schools that serve low income or low achieving students, a sustained focus on improving teacher persistence is long overdue. Put bluntly, preparation programs can help to solve the turnover problem by preparing their graduates to be better teachers. They can train teacher candidates in the kinds of schools where they are likely to teach once they graduate and obtain employment. It may well be that a full year of clinical training is required to produce strong first year teachers. This change, already implemented at places like Arizona State University, ought to be considered even if it pushes some candidates out of the program because they are unwilling or unable to spend a full year learning to teach.

Another program-level step is supporting graduates once they leave: effective induction programs to support program graduates in their first few years of teaching would also help to reduce teacher turnover. A recent paper published by APLU argued, “Like all other beginning professionals, novice teachers are not expert. To become expert they need nurturing and support in their beginning years of teaching. The teacher preparation programs where they initially

³⁴ See the discussion in ³⁴ Gary T. Henry, Fortner, C. Kevin, and Bastian, Kevin C. (2012) “The Effects of Experience and Attrition for Novice High School Science and Mathematics Teachers” *Science*. 335, 1118-21.

³⁵ “Transforming Teacher Education Through Clinical Practice: A National Strategy to Prepare Effective Teachers.” (Washington: National Council for Accreditation of Teacher Education, 2010), p. ii.

developed their skills and where trusting relationships were built should be a part of that support structure for novice teachers.”³⁶

An extension of this philosophy is for programs to work with their graduates in “high impact schools” where they have a critical mass of program graduates in teaching and in school administration. These high impact settings are schools where it might be possible to test, refine, and extend other student outcomes measures like progression and graduation rates, the proportion of students “on track” to graduation, and postsecondary participation rates.

Practical Issues

Some programs do track the persistence rates of their own graduates. But a reliable strategy to acquire data on persistence as a program outcome requires data systems that enable all programs to locate their graduates in the schools and districts where they teach. Thanks to the federally funded State Longitudinal Data System (SLDS) initiative, such systems are becoming more common in the states.³⁷ Data system availability and functionality, however, doesn’t mean that states or programs actually track their graduates and analyze persistence rates.

Making persistence rates a strong operational outcomes indicator will require programs and states to work together to gather and share the data. One of the rare comprehensive efforts to do this has been developed by the Texas-based CREATE Center.³⁸ Obtaining access to the big state databases on licensure and employment—as CREATE does with its member universities—is one approach. For other universities, it might make sense for state agencies to collect and disseminate persistence rate data for preparation programs. And for the many programs that produce a small number of annual graduates, it might be necessary to pool persistence rate results across several years to smooth distortions caused by having a small number of graduates in a single cohort.³⁹

³⁶ Jennifer Presley and Charles Coble, “Seeking Consensus on the Essential Attributes of Quality Mathematics and Science Teacher Preparation Programs.” (Washington: Association of Public and Land-Grant Universities, 2012).

³⁷ These systems are becoming increasingly common in the states, as discussed below and in Appendix A.

³⁸ See <http://createtx.org/content.php?p=36>.

³⁹ Teachers who “stop out” complicate the calculation of teacher persistence rates. When the Illinois Education Research Council looked at five-year persistence rates for programs across that state, it found that about one-third of those who left the profession in their first few years later returned to teaching. Program persistence rates that build a five-year cumulative record for a program cohort (for example, what proportion of 2007 graduates are teaching in 2012) would help to deal with this issue. A dip in persistence rates in the second or third year would be offset later when these graduates return to the classroom four or five years after completing their program.

It is worth saying again here that the use of persistence rates as a program outcome does not mean that preparation programs are solely responsible for teacher turnover. But turnover rates will not improve until producers and employers have incentives to focus on the problem. It seems likely that public confidence in teacher education will be improved when programs take public ownership of this issue.

Meeting state and district needs for teachers

Production of new teachers in high demand fields is a program outcome also highly relevant to the needs and interests of schools and their students. Florida and New York include production of teachers in high-need fields as an explicit focus of Race to the Top. Employment as an outcome measure is part of the Race to the Top strategy for Florida, Massachusetts, New York, Ohio, Rhode Island, and Tennessee.⁴⁰ It's important to note here that Massachusetts, New York, and Rhode Island will use these production and employment numbers as part of beefed-up accountability systems. The other states simply report on them.

Practical Issues

As measurable program outcomes, production and employment outcomes require comprehensive state-level data about program graduates. The state data systems needed for measuring teacher effectiveness as a program outcome—linking K-12 students, their teachers and schools to the programs producing these teachers—would also be necessary to capture information on the production of new teachers in demand fields such as STEM subjects, special education, and ESL.⁴¹

K-12 student perceptions of their teachers

Student surveys as an indicator of teaching quality provide another way to measure program impact on K-12 schools. The Measures of Effective Teaching (MET) project reported in 2010 that student perceptions about instruction were related to teaching effectiveness.⁴² For

⁴⁰ For Florida, this means the production of new teachers in science, mathematics and other STEM subject employed in difficult to staff subjects and schools; New York targets—but doesn't define—"shortage" subject areas. And all the states with program graduate employment indicators focus their attention on high need schools.

⁴¹ As an example, to address teacher needs in Georgia, the University System of Georgia (USG) created a structure for identifying historical and anticipated teacher needs, by licensure area, in all Georgia districts. This was data that USG institutions were encouraged to reference in considering campus teacher education productivity goals.

⁴² "Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project." (Seattle: Bill and Melinda Gates Foundation, 2010), p.7.

example, MET reported that “student perceptions of a given teacher’s strengths and weaknesses are consistent across the groups of students they teach. Moreover, students seem to know effective teaching when they experience it: student perceptions in one class are related to the achievement gains in other classes taught by the same teacher.”⁴³ MET reports that the strongest student perceptions as explanations for learning outcomes are a “teacher’s ability to control a classroom and to challenge students with rigorous work.” School administrators concerned about the classroom management skills of new teachers as well as parents worried that too many teachers have low expectations for their children would understand the meaning of these findings.

MET argues that student perceptions are an “inexpensive way” to construct a teaching quality indicator that can supplement other measures. Of course, the quality of this indicator depends on the instrument used to capture student attitudes. MET employed a survey developed by Ronald Ferguson and his colleagues at the Tripod Project for School Improvement. There are seven dimensions to this instrument: Care, Control, Clarify (teacher explanations, student understanding), Challenge, Captivate (student interest), Confer (teacher questioning), and Consolidate (teacher feedback). A sample item shows the flavor of the survey: “In this class, the teacher expects nothing less than our full effort.” This MET report found statistically significant relationships between some Tripod student responses and teacher value added scores in ELA and mathematics.⁴⁴

The purpose of this discussion is not to argue for a particular instrument for measuring student perceptions of teaching quality. Rather, we propose that CAEP include an indicator that measures student perceptions as a key teacher preparation program outcome.

Practical Issues

Implementing a student perceptions survey as an indicator of program quality in the CAEP accreditation system will require an instrument that meets standards of rigor. Programs may use locally developed instruments for internal purposes, but a CAEP-approved tool will be required for national accreditation. Obtaining survey results will require the cooperation of schools and districts, and there are precedents for this. New York City and the Chicago Public Schools are among the districts that already conduct student surveys on a regular basis. The MET project had

⁴³ Ibid., p. 9.

⁴⁴ Ibid., p. 25-27.

the cooperation of six school districts: Charlotte-Mecklenburg, Dallas ISD, Denver, Hillsborough County (Tampa), Memphis, and New York City.

Distributing, collecting, and analyzing student surveys would be a large logistical task. State data systems could be used to aggregate the data from different schools and link findings to the graduates of specific preparation programs, just as they will have to do for other outcomes measures. The state systems or consortia like the Texas-based CREATE could perform these tasks as well as managing a reporting platform for public dissemination of findings.

Employer and graduate satisfaction with preparation programs

Employer and graduate satisfaction with teacher preparation programs offer two outcome measures that are already being used by a growing number of programs. By themselves, these measures would clearly not be enough to capture the performance or impact of a program. Combined with indicators of student achievement, classroom teaching, and persistence in the profession, however, the feedback of graduates and those who hire them offers a comprehensive picture.

Where these surveys are used, graduates are contacted to find out how well their program prepared them to teach, and some programs solicit similar feedback from principals or other district-based employers of their graduates. Many who talk with schools or school district about teacher hiring hear anecdotes about the graduates of various programs. Some report that a particular provider's graduates are so good in the classroom that they would hire every one of them. Other HR offices or principals are less positive, saying they would never hire someone from such-and-such a program. Districts and schools act on these feelings, but they do not constitute systematic feedback about program or teacher quality.

Practical Issues

Surveys and their response rates must meet standards of quality to yield reliable results. Besides the efforts of individual programs to survey graduates and their employers, there are multi-program or statewide feedback surveys that can be tapped as models. Since 2001, the California State University system has conducted regular surveys of program graduates and their employers, with a common instrument (see <http://tinyurl.com/yetuw85>). Since 1998, the North Carolina State Board of Education has produced an annual IHE Performance Report on program graduates and employer assessment of all state and private teacher preparation programs, with results made available to the public on-line at: <http://www.ncpublicschools.org/ihe/reports/>. In

New York, the Pathways Project implemented follow-up surveys of preparation program graduates and of first- and second-year teachers who had completed programs in the Pathways research initiative (see <http://tinyurl.com/ybgufex>). The Pathways survey findings has contributed rich contextual information about program features, the organization of clinical practice in a variety of preparation programs, and the extent to which preparation of teachers was “coherent” in ways that strengthened the capacity of program graduates to be successful teachers.⁴⁵ Ohio is the only Race to the Top state that plans to implement feedback surveys.

Instruments and survey findings are online for Pathways and the CSU work. In Chicago the Consortium on Chicago School Research conducted surveys of Chicago Public School (CPS) teachers prepared by multiple programs in the area (<http://tinyurl.com/yeabgel>). These surveys were not envisioned as ends in themselves, but as useful sources of information to support research and program improvement. A reliable set of outcomes measures that include survey findings requires data systems that allow all programs to locate their graduates in the districts and schools where they are employed as teachers. It is certainly more feasible for states to collect and disseminate than for 1400 individual programs to develop their own surveys and go off in search of employment data. Moreover, survey quality and response rates must be high enough to allow programs, states, and accreditors to be confident about inferences made from the responses. For feedback measures to be useful to programs, employers, and others the surveys ought to be conducted annually or no less frequently than every other year. Longer intervals between surveys mean that findings will be “stale” as an indicator of program performance and as a program improvement tool.

States Push Outcomes for Accountability Purposes

As states and others work to sharpen a focus on program outcomes, it is important to recognize the extent to which some states are moving to place a stronger emphasis on preparation program outcomes as measures of program quality. The 2012 CCSSO paper discussed above highlights aspects of this trend, as do the efforts in a number of the states to implement Race to the Top commitments that bear on measuring preparation program outcomes. The Chiefs call for transparent reporting systems that make preparation program information

⁴⁵ Research reports and published studies using this information can be accessed at <http://tinyurl.com/5w9ak7>.

available to stakeholders like teacher candidates, policy makers, district leaders, and school hiring officials.⁴⁶

Still needed in the United States is agreement among all states on the quality standards that teacher education will be guided by and judged on to earn national accreditation. CAEP's goals would be aided immeasurably by multi-state agreements on accreditation standards, program outcomes, and licensure tests. With the renewed focus on teacher preparation in states, across the country, and internationally, this is an opportune time for CAEP itself to seize upon this momentum and drive changes in this states through the standards and measures it adopts, and by the expectations it creates for programs that seek initial accreditation or renewal of their accreditation status.⁴⁷ Other professions have managed to adopt and implement the same set of program quality standards and accreditation rules in every state. Nursing, accountancy, engineering, and medicine have uniform state accountability requirements for professional education programs *and for the graduates of these programs*. These policies have been implemented nationwide without undermining professional autonomy, faculty academic freedom, or the principles of federalism—three of the many red herrings raised against national policies by teacher educators.

Putting Outcomes-Focused Teacher Preparation Into Practice

This paper has argued that accreditation of teacher preparation programs should focus on a small set of outcomes. Their value as indicators of program quality is not undermined by the challenges associated with measuring them. Moving as quickly as possible toward adoption of a set of program outcomes—even if some of them might be described today as “ideals” because of measurement and data collection issues—means taking on these challenges as a profession rather than waiting for others to deal with them. The use of teacher tests as measures of candidate knowledge and professional knowledge is the clearest example. Better measures are needed but until they arrive on the scene, CAEP can ratchet up state quality control in significant ways for programs and candidates through the steps we have suggested.

⁴⁶ This is recommendation 9 from “Our Responsibility, Our Promise: Chief State School Officers’ Task Force Report on Transforming Educator Preparation and Entry into the Profession.” (Washington: Council of Chief State School Officers, 2012).

⁴⁷ In particular, CAEP can establish universal standards for all programs in the country with respect to outcomes and measures, and for standards for testing and test passage rates, regardless of the state in which a program happens to be located.

We believe there is much to be gained by limiting the number of program outcomes adopted by the CAEP as the core of its accreditation practices. Operating with a small number of strong and relevant outcomes will make it easier to understand the goals pursued by preparation programs as well as the values and actions of accrediting bodies. Accreditation and state policy can be linked through mutual reliance on a small number of good indicators that are measured consistently (i.e., as near to identical as possible) at *every* institution and program in *every* state with results reported publicly for all programs every year.

The outcomes measures proposed for outcomes-focused teacher preparation can have multiple uses. First, of course, is their value as indicators of preparation program performance when the program is under review for accreditation or reaccreditation. Program faculty and administrators should use the same outcomes to guide ongoing program improvement efforts. States ought to tap these indicators for their work in program approval and accountability, just as the Council of Chief State School Officers states in its 2012 policy paper. And finally, the data collected on program outcomes for accreditation, program improvement, and state accountability will offer a rich trove of evidence to researchers trying to understand what works in the preparation of teachers.

In making the case for these outcomes and measures as the core focus of teacher preparation accreditation practices, the authors recognize that characteristics of the programs themselves play an important role in *how* programs and graduates achieve good results. Here, too, we think less is more in the way of program process standards. But one program feature that deserves careful attention has to do with the quality of clinical settings where teacher candidates learn how to teach. The extent to which a good clinical site contributes to candidate development has an obvious (and measurable) impact on program performance and on graduate teaching quality.⁴⁸ As NCATE's Blue Ribbon Commission argued, "That portion of preparation that is practiced and demonstrated in real schools with real students helps ensure that candidates will be ready for the students with whom they will work and the schools in which they will teach. This is

⁴⁸ A good example of the measurable impact of clinical settings on teaching effectiveness was provided through the research of Matthew Ronfeldt, "Where Should Student Teachers Learn to Teach? Effects of Field Placement School Characteristics on Teacher Retention and Effectiveness." *Education Evaluation and Policy Analysis*, March 2012, 34 (1), p. 3-26.

critically important in preparing teachers to be successful in hard-to-staff, low-performing schools and is useful in all teaching environments.”⁴⁹

The commitment to quality clinical practice for teacher candidates as a dimension of accreditation can be met by keeping in mind that expectations and standards for clinical settings used by preparation programs ought to be crafted so they bear directly on the outcomes indicators. The guiding question should be: How are features of candidate clinical practice and the setting where they learn to teach expected to impact classroom teaching performance as measured by high quality observation instruments; student learning outcomes; measures of candidate and teacher knowledge; persistence in teaching; placement in high need schools and subjects; etc.

Data and Data Systems

Collecting, organizing, and reporting the information needed for outcomes-focused teacher preparation are not easy tasks. Despite some progress over the last five years in improved state data systems, data collection (and use of data) about teacher preparation programs and graduates is quite fragmented and incomplete in 2012. One consequence is the absence of systematic and reliable information about the knowledge, skills, and effectiveness of program graduates outside of a few states (Louisiana, Texas, Florida, Tennessee), a small number of universities that invested their own resources in this work (e.g., New York University, Virginia), and research projects making effective use of access to state datasets (CALDER and Pathways are the best known of these).

Where they exist at all, however, most outcomes indicators other than student achievement are proxies for the concepts, knowledge, and behaviors they claim to measure. And measures of teacher effectiveness are available today for only about one-third of all teachers. Challenges related to the availability and quality of data encompass almost everything about teacher preparation: the characteristics of entering students, their experiences and performance in preparation programs, outcomes such as teaching performance, pupil learning, persistence in teaching, and how teaching context may or may not affect these and other outcomes.

⁴⁹ “Transforming Teacher Education Through Clinical Practice: A National Strategy to Prepare Effective Teachers.” (Washington: National Council for Accreditation of Teacher Education, 2010), p. 27.

This state of affairs is in some ways a reflection of the field itself, where there is still too little agreement on the knowledge and skills that graduates should have and be able to demonstrate in the classroom. Where there is some agreement—on “standards”, “competencies”, and “dispositions”—it resides mostly at an abstract level where the concept is so general as to be often non-observable and unmeasurable in reliable and valid ways. These problems have consequences for accreditation policies and practices, for research about teacher education, for state oversight of preparation programs, and for the efforts of programs to assess their own effectiveness.

What’s Needed Now: An Overview

Data collection systems useful for capturing information about outcomes and available for sophisticated analyses can be tapped for program assessment, policy analysis, and continuous improvement. This kind of system can also help to build an evidence base for what works in teacher preparation. For all this to occur, a robust data collection system must be in place (such as those that Race to the Top states are building or adapting) to generate mainly aggregate measures of preparation program outcomes from individual-level data, or from datasets with links between files containing information about students, teachers, schools, and preparation programs. Data elements, data collection protocols, and management of the system(s) by multiple parties⁵⁰ have to be configured to produce accurate data.

To understand outcomes-focused teacher preparation, these are the data system linkages that matter most:

- School link to teachers
- School link to pupils
- Classroom-level data: classes, teachers, and pupils
- Pupil individual identifier
- Pupil demographics
- Pupil test data
- Pupil link to teachers
- Pupil link to classes
- State certification data for teachers
- Teacher employment records

⁵⁰ Such as universities and university systems, state agencies, schools and school districts, and federal government (IPEDS, Core of Common Data, other NCES resources).

As just one example, information on individual schools and employed teachers is necessary to calculate persistence rates in teaching for program graduates. The National Commission on Teaching and America’s Future (NCTAF) described three types of teacher turnover.⁵¹ It is not easy to determine whether a specific individual teacher has left teaching entirely, but data about teacher employment at school and district levels are needed to calculate and report the most widely used measures of persistence and turnover.

Given the wide range of information needed on teachers, students, and schools, a system that meets these conditions will probably be a compatible set of independent databases maintained by different parties and linked through common identifiers. Examples already exist, such as the one developed through North Carolina’s Education Research Data Center (http://www.childandfamilypolicy.duke.edu/project_detail.php?id=74). The Data Quality Campaign’s “Essential Elements” and “10 state actions to ensure effective data use” provide an overview of how comprehensive data systems need to work if they are to be useful (see <http://www.dataqualitycampaign.org/build/elements> and <http://www.dataqualitycampaign.org/build/actions>).

Data systems with important cross-file linkages: Examples

Florida Data Warehouse All necessary linkages are built in to the design of the system. Researchers and others interested in the program outcomes are not required to create cross-file linkages. As with North Carolina below, this is the ideal system design.

North Carolina Education Research Data Center The linkages needed for comprehensive data about candidates, teachers, schools, and pupils are built in to the design of the system. Thus, Florida and North Carolina—one a state system, the other a project system built with state cooperation—are the ideal systems in this regard. Researchers and other users are not forced to match files themselves, because the design of these two systems includes use of common identifiers across files. In other systems profiled here, users are required to develop and implement strategies to make these cross-file matches.

⁵¹ NCTAF defined teacher turnover in terms of: **Leavers, or** teachers employed in a classroom-teaching role in a school in Year 1 and not employed as classroom teachers *in any district* in Year 2; **Within-District Movers**, teachers employed in a classroom teaching role in a school in Year 1 who are employed as classroom teachers at a different school *in the same district* in Year 2, or “cross-school, within-district movers;” and **Cross-District Movers, who are** teachers employed in a classroom teaching role in a school in Year 1 who are employed as classroom teachers at a different school *and in a different district* in Year 2.

Texas PK-16 System This system—known as the Texas PK-16 Public Education Information Resource (<http://texaseducationinfo.org>), has very comprehensive data resources but it is not clear whether all of the cross-file linkages exist. Reports from researchers or universities seeking to use this data raise questions about ease of access, noting obstacles placed in the way by the Texas Education Agency (TEA).

Utah Data Warehouse This is a very comprehensive system of K-12 data, but it is not clear whether links exist to files on teacher preparation institutions. (<http://www.usoe.k12.ut.us/upass/>).

Louisiana As with Utah and Texas, cross-file linkages are not uniformly built in to the design of the data warehouse (www.doe.state.la.us/lde/pair/638.html). As with Utah, the system and the access site in Louisiana continues under development.

New York Pathways The Pathways Project has created all of the linkages needed to conduct analyses relevant to the study questions (www.teacherpolicyresearch.org). However, project leaders were forced to assemble the data systems and make the cross-file linkages themselves because of the poor quality of data systems in New York State and New York City.

California State University The CSU system has cross-file linkages to individual level data for college students who enrolled in or completed teacher education programs offered by one of the System's campuses (<http://calstate.edu/teacherquality>). Links to files about schools, teachers, and pupils depend on partnerships with individual school districts. These arrangements have been assembled through negotiations and related CSU project development activities.

Given how teacher preparation programs actually work in practice, the best system configuration for teacher education would use interstate data system linkages to cope with mobility of teacher candidates and program graduates across state lines. As Secretary Duncan has reported to Congress, for instance, 20 percent of initial teaching licenses in 32 states were granted to new teachers prepared for the classroom in a different state from the one granting the license to teach. In another 12 states and the District of Columbia, programs in other states prepared 40% of initially certified teachers.⁵²

The optimal system—a comprehensive data system at the national level—is highly unlikely ever to be available. Efforts to construct such a system in the last decade were blocked by a coalition of political conservatives and independent higher education institutions. Within specific states, universities, or school systems, missing pieces include large chunks of relevant data,

⁵² See U.S. Department of Education, “The Secretary’s Sixth Annual Report on Teacher Quality,” page 23. Retrieved from <https://title2.ed.gov/secReport11.asp>.

ability to link datasets with common identifiers, barriers constructed at every level in the name of “privacy”, and technical problems with hardware, software, or staffing capacity.

Gaps in data system components and dataset linkages are gradually being bridged. But they still exist. In spite of these challenges, some information needed for solid answers about preparation program outcomes already exists. Examples include the Pathways Project, AIR’s Center for the Analysis of Longitudinal Data for Education Research (CALDER), the Texas CREATE initiative, the California State University system, and the data systems behind publications on preparation program effectiveness from states like Louisiana, Florida, Tennessee, and Texas. Appendix A discusses the components and details of ideal and current data systems in more detail.

Accreditation and State Teacher Quality Policy: Opportunities for Change

States use program approval criteria to decide which institutions and other organizations are sanctioned to train teachers. *These decisions establish the “universe” of preparation programs eligible for national accreditation.* A key implication of this fact is that state actions are central to the quality and viability of program accreditation. It is this connection that leads us to touch on the relevance of outcomes, data, and measurement to the states and to CAEP.

State program oversight policies are paired with administrative processes like program reporting cycles to state agencies, periodic reviews conducted by the state, and campus visits by review teams appointed by the state. While most states may not use pupil learning results or other outcome measures to evaluate teacher preparation programs,⁵³ they do put extensive time and energy into developing and administering program approval and oversight practices. These are the responsibility of a state department of education or of a “professional standards board” with jurisdiction for teacher education program approval and teacher licensure. States manage these activities through a system of policy directives, teaching standards, campus visits, voluminous documentation requirements, and an appeals process when programs are unhappy with the results.

⁵³ With the limited exceptions of Louisiana, Florida, and Texas for some outcomes, and the possible addition of five of the states supported by federal Race to the Top grants (DC, Maryland, Massachusetts, New York, Rhode Island) although actual accountability systems have yet to be implemented in these states.

The details of all this vary greatly by state, but it is safe to say that teacher preparation program accountability is far more decentralized in the United States than in most other developed (or OECD) countries. It is also more fragmented than certain other types of professional education in the United States.⁵⁴

Some states complement their state regulations with national accreditation, a voluntary process for assuring quality control. State partner arrangements with NCATE often include joint program reviews that drive state approval *and* accreditation decisions. The overlap between accreditor and state agency program oversight can be confusing to the public and to elected and appointed state policymakers, and it changes regularly as states move in and out of partnerships and agreements.

We suggest below ways that accreditation and state policy can work together in the service of high standards, effective quality control for preparation programs and their graduates, and to advance the educational success of K-12 students. CAEP can move the states forward to meet their obligations to students and taxpayers. There are some useful lessons from other professions where accreditation and state policy reinforce each other to support similar goals.

Comparisons with other professions

For the most part, state teacher education policy has achieved little of what is evident from the experiences of other fields. For these professions—nursing, engineering, accountancy, and medicine are examples cited earlier—state policy and accreditation have created and sustained *national consensus* about a set of values that shape essential activities in each field: who enters the profession, how they are trained (and at which programs), how they practice, and how programs and practitioners are regulated. As just one example, entry standards into professional practice are established and enforced by these professions through regulatory bodies such as licensing boards. For this to work, governments (in the United States this means *state* governments) must accept the norms of the profession itself with respect to which individuals “deserve” entry. In some professions other than teaching, these are *national* norms.

Those who study occupations and professions argue that recognized professions evolve from a community with shared norms, training, working practices, and regulatory mechanisms.⁵⁵

⁵⁴ Some examples of more coherent oversight even within a federal structure are medicine, nursing, engineering, and accountancy.

⁵⁵ James Begun and Ronald Lippincott, “Strategic Adaptation in the Health Professions” (San Francisco: Josey-Bass, 1993), p.31.

When the professional community is strong enough, the same values can be found at the heart of training programs, program accreditation and approval processes, licensure policies, and professional practice. The result is professional legitimacy.⁵⁶

For accreditation to be effective in promoting and sustaining quality control and strong outcomes, decision-makers need to recognize how education, accreditation, and licensing are linked as part of a policy continuum. What some sociologists and economists call credentialing is a key source of professional legitimacy; its components are education, accreditation, and licensure. Credentialing includes receipt of degrees or certificates by individual members of the profession, licensure by the state, and accreditation of education programs by some external organization.

One consequence of professional legitimacy is success in using government regulatory power to establish and enforce entry standards. Thus, individuals seeking to become doctors, engineers, or nurses can only be licensed if they complete an *accredited* program of professional education and pass one or more tests. For these professions, every state recognizes the same set of tests and passing cut scores. Program accreditation is a prerequisite to professional licensure in these fields because all states recognize and accept the sanctioning power of a non-profit entity as a gatekeeper.

We do not argue here that teaching licenses should be granted only to the graduates of accredited teacher education programs. However, the disconnects between program accreditation and state teacher quality policy create significant problems for preparation program accrediting bodies and they undermine claims to professional status by the field itself. One reason for the difference between teacher preparation and other forms of professional education in this regard is that establishing and maintaining authority through the professional community (i.e., accreditation) cannot happen “unless its members agree, first, on criteria for belonging to the profession, and second, on what its rules and standards ought to be.”⁵⁷

If program accreditation is part of the process through which a profession sets and applies its own rules, then to be effective inside and outside the field, it must be rooted in rational and scientific ideas. Those outside the field must see accreditation standards as effective mechanisms

⁵⁶ Begun and Lippincott; Paul Starr, “The Social Transformation of American Medicine.” (New York: Basic Books, 1982).

⁵⁷ Starr, p.80.

of quality control.⁵⁸ In contrast with some other fields, however, the claims to professional authority asserted by teacher preparation accreditation and licensure do not rest on agreements within the field on rules and standards. Nor do they currently have necessary links with scientific knowledge or with results that benefit the public (e.g., student outcomes).⁵⁹

CAEP and the states

The absence of widespread agreement within teacher preparation about rules and standards as well as the weak reliance on outcomes and results to approve or accredit programs and license their graduates are defining characteristics of teaching and teacher preparation in the United States. This sets up a dilemma for CAEP in how it relates to the states. The dilemma is made worse because state teacher quality policies generally have not promoted high standards for preparation programs or for entry into teaching. On the other hand, since some states now appear ready to raise the bar on program quality and licensure for teachers, CAEP can help to improve weak state policies in ways that begin to tighten the connections among professional education, accreditation, and licensure. This would be a very big step to ensure that all three processes—which are intertwined in all forms of professional education—are driven by similar values and desired outcomes.⁶⁰

Taking a stand: CAEP policies

CAEP can and should take some steps now in the hope that states will follow its lead. For example, the Commission ought to set and enforce high standards, driven by transparent evidence-based performance indicators of the kind discussed in this paper. CAEP should do this independent of what any state’s teacher quality regulations “settle for”. This means that as a national accreditor CAEP ought to make its requirements nationwide with respect to (a) outcomes indicators that programs have to adopt and report on to be considered for accreditation; (b) acceptable levels of performance on these indicators to achieve and maintain program accreditation; (c) the licensure tests and test passing scores that it accepts as evidence of quality from any program that seeks accreditation, regardless of state; and (d) the licensure test passing rates that it sets as a floor for program accreditation. We argue elsewhere in this paper that

⁵⁸ Begun and Lippincott; Starr.

⁵⁹ Begun and Lippincott.

⁶⁰ Parts of this discussion draw on previously published work by one of the TPA authors (Edward Crowe, “Teaching as a Profession: A Bridge Too Far?” In Marilyn Cochran-Smith et al., “Handbook of Research on Teacher Education.” New York: Routledge, 2008).

CAEP should lead the way through its accreditation practices towards a significant reduction in the number of teacher tests recognized as relevant to program performance, should take a stand by only recognizing test passing cut scores set at the 75% percentile, and should require a program-wide passing rate of at least 80% on teacher tests whose cut score is set at the 75% percentile.⁶¹

We argue for a passing score standard well above the mean for two reasons. As the CCSSO says in its recent policy paper, “Today, however, we are asking licensure assessments to do more, to ensure a certain standard of educator quality and to be based on indicators correlated with readiness to enter a classroom or a school so we can make better informed decisions of who gets into the profession. Current reform efforts are focused on these new expectations of performance — Can the candidate actually do the job? — and higher standards of rigor — Are educators effective?”⁶² Secondly, establishing a cut score at the mean of all test takers—which would be a significant step up for many states—communicates to the public that we are looking for “average” individuals to become teachers. But average is not good enough for our children, nor is it good enough for a field that seeks to improve its status and its public reputation.

If alternative cut scores to the 75th percentile are adopted by CAEP, such as setting the score one standard deviation above the mean for all national test-takers (which would be the 68th percentile),⁶³ CAEP should follow the same principle discussed above: CAEP’s standard should apply to all programs seeking accreditation, regardless of the passing scores adopted by the state in which the program is located. And, as noted previously, if a state sets its own cut scores *higher* than the CAEP standard, CAEP should employ that higher standard for all programs in that state seeking accreditation.

In recent weeks, both the Council of Chief State School Officers (CCSSO) and the American Federation of Teachers have called on teacher preparation programs to implement selective admissions practices. This is not a new issue but CAEP can help the country make progress in the quest to recruit, enroll, and prepare academically strong teacher candidates. In particular, the resistance of individual preparation programs to selectivity in admissions and exit criteria can be countered for accredited programs (and those who seek accreditation) by setting

⁶¹ With no gamesmanship on definitions of who is a “program completer.”

⁶² CCSSO, “Our Responsibility, Our Promise: Transforming Educator Preparation and Entry into the Profession,” p. 14.

⁶³ See the discussion on pages 15-16 in this paper.

teacher test cut scores high (at the 75th percentile or at least one standard deviation above the national mean score for all test-takers) and by requiring test passing rates of 80% or better. High institutional failure rates, when published openly, will drive institutional change and may also reduce the oversupply of program graduates in fields like elementary education.⁶⁴

The foregoing recommendations represent actions that CAEP should take as it builds a credible system of preparation program accreditation. As long as many states persist in their own low standards, CAEP has to find a way to disentangle itself from these policies.

Working with the states

Beyond these actions, however, there are specific steps that CAEP can take in its relationship with the states. There is much to be gained by working with state policymakers to achieve these goals over the next few years:

- Set and enforce high standards, driven by transparent evidence-based performance indicators.
- Alignment of state data systems and preparation program outcomes indicators so that teacher education programs in every state, should they choose to do so, can generate reliable and transparent reports on all of the indicators adopted as part of CAEP's outcomes-focused accreditation process.
- To this end, CAEP should consider working more closely with states to improve data collection and reporting, perhaps even by pursuing a strategy of dividing up the responsibility for data collection and reporting between institutions, states, and districts – both for accountability and program improvement purposes.
- Revision of state program approval and reapproval processes to focus them on the same set of measurable outcomes that CAEP will use to make accreditation and reaccreditation decisions. Here again, the CCSSO recommendations provide a good starting point for specific actions by CAEP and by states willing to move in this direction.
- Use data on these outcomes indicators collected by and about preparation programs to launch in-state studies of program features that yield successful K-12 outcomes. Models for these

⁶⁴ APLU has also called for “a set of rigorous standards and performance assessments that must be taken by all candidates seeking teacher licensure in science and mathematics (and other areas as well).” See Jennifer B. Presley and Charles R. Coble (2012). “Seeking consensus on the essential attributes of quality mathematics and science teacher preparation programs”. APLU/SMTI, Paper 6. Washington, DC: Association of Public and Land-grant Universities. <http://www.aplu.org/document.doc?id=4098>.

studies and the organizations that produced them can be found in Louisiana, North Carolina, New York, and Washington State.

- Reduce duplication and needless paperwork in the program approval and reapproval processes by focusing on the program outcomes that matter as well as preparation program characteristics⁶⁵ that can be tied directly to these outcomes.
- Foster a movement that results in real alignment of accreditation, program approval, and licensure policies along the lines developed in other professions. Nursing education may be an excellent model to study and adapt because of the work over the past 30 years to achieve this very goal.

Both CAEP and the states would benefit from a set of policy and practice changes along the lines discussed here. Current opportunities and pressures for reform can be leveraged to improve preparation program quality, strengthen the ability of program graduates to be effective teachers, and continue the difficult work of raising the status of teaching and teacher preparation in the United States. Failure to act in decisive ways may doom national program accreditation to irrelevance, and it will also hurt the students and schools who represent the future of our country.

⁶⁵ For example, recruitment and selection processes for candidates into teacher preparation programs are significant program characteristics. Others include features of the ways in which programs organize their clinical components (length of time for student or intern placement in practice teaching; nature and details of the observation and feedback system used to promote teaching skills development in candidates; how schools are selected, staffed by the program, and supported as clinical sites), linkages between coursework and fieldwork such as those explored by Grossman and colleagues in “Constructing Coherence: Structural Predictors of Perceptions of Coherence in NYC Teacher Education Programs.” *Journal of Teacher Education* (2008) 59.

Teacher Preparation Analytics, LLC

Founders and Partners

Dr. Michael Allen is a Senior Research Consultant for CNA Education and a Senior Consultant to the Science and Mathematics Teaching Imperative of the Association of Public and Land-grant Universities. His work has focused primarily on math and science education, teacher education, teacher supply and mobility, and teacher licensure and assessment. In addition to recent reports on early mathematics and reading education, he directed a two-year NSF-funded study that produced a report and resource guide to help state policymakers and university officials better assess the state-level need and supply pipeline for science and math teachers. He has done other work on the indicators of quality for teacher preparation program, and as a senior program officer at the National Research Council he directed the Committee on Teacher Preparation Programs. He served for seven years at Education Commission of the States as a senior policy analyst and director of the teacher quality program. While at ECS, Michael wrote many policy and research-related publications, including two major reports entitled *Eight Questions on Teacher Preparation: What Does the Research Say?* And *Eight Questions on Teacher Recruitment and Retention: What Does the Research Say?* A former professor of philosophy and ethics, he has also published articles on epistemology and a book on nonprofit ethics. He earned a Ph.D. and A.M. in philosophy from Boston University and an M.Ed. in research methods from Charles Sturt University, in Australia.

Dr. Charles Coble is regarded as a national expert on teacher education programs and teacher development. He is a former professor of science education and dean of 13 years at the nationally award-winning school of education at East Carolina University, Greenville, North Carolina. He served for six years as the vice president of university-school programs for the 16-campus University of North Carolina (UNC). In that capacity, Charles led the development of the University School Teacher Education Partnerships in all UNC teacher preparation institutions, and organized the UNC Center for School Leadership Development – a confederation of high-quality programs for the initial and continuing preparation of teachers and administrators. Charles also led the design and construction of an \$11 million state-of-the-art professional development center facility that also houses the James B. Hunt Jr. Institute for Educational Leadership and Policy. He has directed over \$12 million in grants and contracts and is the author or co-author of 10 books and over 70 published articles. Charles served as vice president of policy studies and programs at the Education Commission of the States, and he is a founder and partner of The Third Mile Group, a limited liability company that focuses on education and social policy. A native of North Carolina, he earned his undergraduate, masters and doctoral degrees from the University of North Carolina-Chapel Hill.

Dr. Edward Crowe is Senior Adviser for the Teaching Fellows Program at the Woodrow Wilson National Fellowship Foundation. Over the last ten years, he has provided consulting services on teacher quality and K-16 policy issues for the State Higher Education Executive Officers and public higher education systems of Ohio, Pennsylvania, and Wisconsin. For the National Commission on Teaching and America's Future, he did research on the cost of teacher turnover and other teacher preparation issues, and he was as an adviser to the Hunter Foundation of Scotland and to the Scottish National Executive. Ed has served on the Advisory Council for the Texas Center for Research, Evaluation and Advancement of Teacher Education; as a member of the national advisory panel for the Ohio Teacher Quality Partnership; as an evaluator for the New York City Partnership for Teacher Excellence; and with the Committee on Teacher Preparation Programs for the National Research Council. He was the first director of the Title II Teacher Quality Enhancement Program for the U.S. Department of Education. Ed has extensive experience in state higher education policy, having served with the Office of the President of the University of North Carolina system, with the education coordinating board in Arkansas, and as senior manager of a statewide math and science education reform project funded by the National Science Foundation. He is a graduate of Boston College and holds masters and doctoral degrees in political science from the University of North Carolina-Chapel Hill.

Appendix A

Organization and Features of Data Systems to Support Program Outcomes

Comprehensive data systems to support accreditation-driven program reporting will pull in pieces of information from multiple sources. In many states, component parts of what should be a unified system are data elements collected and managed by different agencies or institutions. Connecting the pieces requires the ability to link files successfully through a unique individual identifier for each teacher candidate and teacher. As the Data Quality Campaign notes, the purpose of this identifier is to link individuals across databases and across years. The latest information is that 44 states now have a statewide teacher identifier with a teacher-student match.⁶⁶ Fewer states have unique identifiers for teacher candidates, but if the outcomes focus pursued by the CAEP Commission is about *program graduates*, this is less of a problem. As noted earlier, current strong data systems include those in Florida, Texas, and North Carolina, as well as the data system constructed by the Pathways project research team in New York. North Carolina's data system is widely used by researchers conducting dozens of teacher effectiveness studies, built at Duke University through multi-agency cooperative agreements.

Among the chunks of data that matter for productive use of these systems, extensive demographic background information on candidates and teachers is important to understand program recruitment practices, for teacher supply and demand studies, and for assessments of program outcomes. Most states have the capacity to collect, aggregate, and report this information for employed or certified teachers. Far fewer can do the same for teacher candidates. Those with this capacity currently are the Florida Data Warehouse as well as the North Carolina, Texas PK-16, and New York Pathways data systems.

Data about Teachers

SHEEO (2003), the Pathways Project, Goldhaber (2003) and others discuss information needed about teachers in order to study teacher preparation programs, pathways, outcomes, and persistence in teaching. This includes data about individual teachers (from licensure, schools,

⁶⁶ According to the DQC, states without this capacity are the District of Columbia, Delaware, Connecticut, Massachusetts, Maryland, Puerto Rico, and Rhode Island. Five of the seven (all except CT and RI) are Race to the Top states that committed to data system upgrades in exchange for large federal grants.

retirement system, and unemployment insurance databases), where they are working, their teaching assignments (grade and subject), their certification and endorsement, salaries and benefits, and their employment status over a multi-year period (persistence in teaching at one school, moving to another school or district, leaving the profession of teaching). It is also crucial to have teacher files that link individual teachers to their pupils for analyses of “teacher effects” on pupil achievement. To obtain the necessary data, SHEEO and ECS indicate these databases should be accessed, improved, and linked:

- Completer database
- Teaching assignment database
- Employment database
- Licensure/certification database
- Student (higher education) database
- Schools database
- Retirement system database, and
- Unemployment insurance database

The names given to these data system components vary across systems and states, but the contents are generally similar. As discussed earlier, most current state and teacher preparation research link teachers, schools, and students. More problematic is how many are able to link individual teachers to their preparation pathway. This linkage is accomplished through a unique individual level identifier shared with other systems or cross-walked with identifier(s) used by the other systems. Through the individual identifier (SSN or some other unique number) in *teacher* files, teachers are linked to their academic records as teacher candidates. Similarly, individual identifiers in the *teacher* files are the mechanism for linking teachers to the *schools* where they teach and to the *pupils* in their classes.

For example, in their research paper on “The Impact of Teacher Preparation on Student Learning in North Carolina Public Schools,” Gary Henry and his colleagues at UNC-Chapel Hill explain how they merged K-12 class rosters, student performance data, and school characteristics with information about teacher education licensure. Dan Goldhaber and Stephanie Little, in “The Gateway to the Profession: Assessing Teacher Preparation Programs Based on Student Achievement,” describe a similar process of linking separate datasets in Washington State.⁶⁷

⁶⁷ Gary Henry, Charles Thompson, C. Kevin Fortner, Kelly Purtell, Rebecca Zulli, and David Kershaw. “Technical Report: The Impact of Teacher Preparation on Student Learning in North Carolina Public Schools.” (Chapel Hill: Carolina Institute for Public Policy, 2010), p. 7-8; Dan Goldhaber and Stephanie

These linked databases are needed to answer questions about persistence in teaching, and about the extent to which program graduates promote pupil-learning gains. The individual teacher identifier is the linkage mechanism from *teacher* files to *school* files, from *teacher* files to *candidate* files, and from *teacher* files to retirement system and unemployment insurance databases. Calculating persistence rates and employment outcomes requires information about teachers from other state records on employment such as the unemployment insurance files.

Links to classroom

Studies of teacher effects must link a variety of data elements to information about the teacher's classroom. The data systems most in use by teacher effectiveness researchers make that connection. The North Carolina Education Research Data Center offers the clearest explanation of this linkage in a description of its files: "Across all files, school codes have been standardized so that researchers can link the student and teacher data sets to every school file. These identifiers permit studying such factors as the impact of school size, demographic composition, or percent of classes that are honors or remedial on teacher quality and student achievement."

Links to pupils taught

The data system in question must support connections between teachers, schools, and the individual pupils taught by a particular teacher. Published studies and reports using the data systems in Florida, Louisiana, Texas, North Carolina, and New York document the existence and utility of these links. These studies also point out the pitfalls of many data systems around the country: the teacher-student test score link may connect exam proctors to students who completed a particular state test, failing to link the teacher(s) responsible for academic outcomes that are measured by the test.

Information about Teacher Candidates

Pre-college test scores

Records of SAT, ACT and similar tests taken as part of the college admissions process are relevant to questions about program admissions practices, selectivity, and student quality. The Florida Data Warehouse, and the Texas PK-16 system capture this data in ways that others can emulate. University databases are typical sources for this information. The Lumina

Little, "The Gateway to the Profession: Assessing Teacher Preparation Programs Based on Student Achievement: Working Paper 65." (Washington: National Center for Analysis of Longitudinal Data in Education Research, 2012), p. 9-11.

Foundation analysis of state higher education data systems indicates that college test score data are widely available (see *Illuminations: Critical Connections*, the August 2010 report published by Lumina (www.luminafoundation.org)).

Courses and Grades

The Florida Data Warehouse has the best structure and resources for this information, aided by state-mandated common course numbering across all public two- and four-year institutions in the state. This means that researchers can study a pool of Florida teacher candidates from multiple institutions or successive years, with greater confidence that they are capturing the same general exposure to academic content.

The Florida Data Warehouse, Louisiana, the Texas PK-16 system, and the CSU system appear to include relevant data. While course grades have limited value as useful measures of anything relevant to program quality or graduate performance in comparative studies, they might be helpful measures at the tails of GPA or course grade distributions. Clearly defined learning outcomes and common high-quality assessment instruments are essential to understand course, program, and institutional outcomes. Useful measures seem decades away.

Clinical placements

Information about the clinical placements of teacher candidates is absent from all state data systems. Where it exists elsewhere this information is kept by individual programs or colleges of education, or collected by them for specific purposes (there are good examples at Arizona State, New York University, and the University of Washington). Detailed information on the location, duration, supervision, faculty involvement, and other aspects of candidate clinical placements are necessary for meaningful analyses of their impact on candidate development, program quality, or program outcomes. Pathways studies on program coherence, quality of clinical sites, and characteristics of preparation programs all point to the importance of these data.⁶⁸ The fact of so few strong studies about these program characteristics shows the difficulty of obtaining high quality information in these areas.

⁶⁸ See: Pam Grossman et al., “Constructing Coherence: Structural Predictors of Perceptions of Coherence in NYC Teacher Education Programs.” *Journal of Teacher Education* (2008) 59, p. 273-87; Donald Boyd, Pamela Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff, “Teacher Preparation and Student Achievement.” *Educational Evaluation and Policy Analysis*, (2009) 31 (4), p. 416-40; Matthew Ronfeldt, “Where Should Student Teachers Learn to Teach? Effects of Field Placement School

Clinical grades and evaluations

Most information about candidate performance in clinical settings gathered by programs and by states is based on low quality assessment instruments and practices within programs. There are very few multi-program data sources on teacher performance in the classroom. Comparisons across programs are largely meaningless. In some states, the edTPA pilot may be a source of information as candidates exit their programs. For reliable quality measures of candidate clinical performance, investment in high quality clinical assessment protocols will be needed, as well as faculty training. Only then will it be possible to draw any useful conclusions about clinical performance of teacher candidates.

Degree, Completion Year, and Institution

Calculating and reporting program outcomes requires accurate information about degrees awarded to teachers, when they completed a program, and which institution awarded the degree which led to state licensure or certification. As researchers have discovered, however, it can be difficult to determine *which degree* obtained by a teacher is the one that functioned as the gateway into licensure and professional practice. The proliferation of alternative certification programs—many at universities with traditional programs—clouds the picture even more. The solution may lie in more reliable information about coursework and clinical experiences, enabling researchers to “go behind” a particular degree (e.g., bachelor’s or master’s) to the academic and clinical preparation that led to award of the degree. This step requires common course outcomes and measures of attainment for teacher candidates enrolled in the courses.

Most state systems provide reliable access to information about year of completion for candidates. *What program or degree* was completed in a particular year is a more complicated issue from the standpoint of licensure or certification, as discussed below. Thanks to the federal Integrated Postsecondary Education Data System (IPEDS), information about which institution awarded a particular degree is generally and reliably available. The Lumina Foundation analysis of state unit record data systems for higher education reinforces this point.

Characteristics on Teacher Retention and Effectiveness.” *Education Evaluation and Policy Analysis*, March 2012, 34 (1), p. 3-26.

Licensure and certification

The best data systems capture information at the individual teacher level about test scores, dates tested, and certification areas covered by the tests. Few such systems appear to exist today, but Race to the Top states and some others are building out their capacity in this area by creating linkages across disparate data sets. Chronological information on testing dates or attaining certification matters for several reasons: delayed entry into teaching after completing a program is fairly commonplace. It is one of the factors complicating supply and demand studies of teacher manpower. Secondly, the number of times a given candidate must take a test in order to meet the state cut score is relevant to questions of candidate and program quality. Some states and programs have taken intentional steps to obscure this fact, while professional programs and licensing boards in law, for instance, routinely report pass rates by individual separately for first-time takers and multiple test-takers. Acceptable quality in this data element should include full information on all dates, tests, and certifications for each candidate.

Current P-12 schools employing program graduates

Florida, North Carolina, and New York have assembled the most comprehensive links between individual teachers and the schools where they teach. Others are not far behind. Essential for analyses of teacher job placement, persistence in teaching, and pupil learning outcomes are these data elements:

Subject(s) taught

Essential information to capture teacher supply and demand flows, as well as efforts to obtain matches between preparation patterns and actual teaching, the data element is captured adequately by state systems such as Florida, Louisiana, Utah, North Carolina and New York. It would appear, however, that Florida and North Carolina have an easier time matching files to obtain the link between teachers and their subject areas.

Course(s) taught

While many state data systems include the courses actually taught by individual teachers—essential for understanding the impact of preparation programs and for matching teachers with pupils—Florida and North Carolina appear to have the best-developed design and collection mechanism for this data element.

Teaching experience

State and district data systems now routinely include a data element on teaching experience, but potential users of the information should proceed with caution. Only by delving into a specific data system does it become clear how definitions and external factors affect the quality of data on teaching experience. Need is information on total teaching experience, teaching experience in the current state of residence, and teaching experience in the school where a teacher is currently employed. For teachers who transfer districts or move from one state to another, total teaching experience *recognized for placement on hiring grids* is capped at about five years. This distorts the total experience and total state experience figures for any teacher who has moved. In addition, some schools negotiate and recognize additional years of experience (beyond actual experience) for teachers in shortage fields or high demand areas. Finally, some districts cap the total experience for all teachers at some figure determined through collective bargaining.⁶⁹

Schools and pupils: contextual data

The context for teaching and learning is an important component of understanding preparation program outcomes. States, districts, and the federal government all collect and archive information about individual schools. This includes demographic information about students, the school's size, grade levels, achievement performance, and other salient characteristics. The Core of Common Data (CCD) collected and maintained by the National Center for Education Statistics (NCES) includes many of the variables apart from learning outcomes. There are some questions about the accuracy and completeness of this CCD information (which can only be as good as the data submitted to NCES by the states). However, a growing number of states are improving their efforts to collect, clean, and maintain comprehensive data about every school. The Data Quality Campaign continues its work to monitor and assist the states.

To implement the CAEP Commission's charge, it will be necessary to connect a number of data sets discussed in this paper: linking data about candidates, teachers, schools, and pupils. The Data Quality Campaign's work has been intended to prod states to design these linkages into

⁶⁹ Milwaukee, for instance, has capped total recognized teaching experience at 16 years for all teachers in the district. District—and sometimes state—personnel files record experience according to these policies, not actual experience.

comprehensive data systems. The Institute for Education Sciences (IES), through its State Longitudinal Data Systems grant program, aims to create the linkages across data system segments; however, the IES effort has a major flaw because grants are made to state K-12 agencies and largely ignore the central role of higher education.

Finally, all large-scale studies of pupil learning include a wide array of information about K-12 pupils—including test scores over several consecutive years, demographic information, and similar data about other students in the same classroom and school (used as study controls for student and school effects). As with teachers, files that can link pupils with their teachers are necessary to conduct studies on pupil learning as a teacher preparation program outcome. Significant progress has been made over the last few years in developing state-level data systems capable of managing and linking these datasets.