MEASURING MASTERY

BEST PRACTICES FOR ASSESSMENT IN COMPETENCY-BASED EDUCATION

Katie Larsen McClarty
and Matthew N. Gaertner

Center for College & Career Success, Pearson

April 2015

AEI Series on Competency-Based Higher Education

CENTER ON HIGHER EDUCATION REFORM

AMERICAN ENTERPRISE INSTITUTE

Foreword

Rising tuition prices and finite public budgets have spawned a lively policy debate about innovation in higher education. In particular, competency-based models have garnered a lot of attention from policymakers, reformers, and funders. Unlike online college courses, which often leave the basic semesterlong structure intact, competency-based models award credit based on student learning, not time spent in class. As soon as a student can prove mastery of a particular set of competencies, he or she is free to move on to the next set. A number of institutions are currently engaged in these efforts, including Western Governors University, Excelsior College, Northern Arizona University, and the University of Wisconsin's UW Flexible Option.

The competency-based model presents opportunities for improvement on two dimensions: first, it allows students to move at their own pace, perhaps shortening the time to complete a degree, and second, competencies can provide a clearer signal of what graduates know and are able to do. Yet for all the enthusiasm that surrounds competency-based approaches, a number of fundamental questions remain: What kinds of students are likely to choose competency-based programs? How do students in these programs fare in terms of persistence, completion, and labor market outcomes? Are these programs more affordable than traditional degrees? What does the regulatory environment look like for competency-based providers? Do employers value the credential?

Despite increasing attention being paid to the potential of competency-based education, researchers and policymakers still have few answers to these questions. To provide some early insight, AEI's Center on Higher Education Reform has commissioned a series of papers that examine various aspects of competency-based education. In the third paper of the series, Katie Larsen McClarty and Matthew N. Gaertner of Pearson Education introduce a set of best practices for high-stakes competency-based education assessment, detailing how providers can work to validate their assessments and establish performance levels that map to real-world mastery.

As always, our goal is not to come up with a verdict as to whether this innovation is good or bad, but to provide a look under the hood that is useful to policymakers and other observers. I hope you find it useful, and stay tuned for more.

—Andrew P. Kelly
Resident Scholar in Education Policy Studies
Director, Center on Higher Education Reform
American Enterprise Institute

Executive Summary

Competency-based education (CBE) programs are growing in popularity as an alternative path to a postsecondary degree. Freed from the seat-time constraints of traditional higher education programs, CBE students can progress at their own pace and complete their postsecondary education having gained relevant and demonstrable skills. The CBE model has proven particularly attractive for nontraditional students juggling work and family commitments that make conventional higher education class schedules unrealistic. But the long-term viability of CBE programs hinges on the credibility of these programs' credentials in the eyes of employers. That credibility, in turn, depends on the quality of the assessments CBE programs use to decide who earns a credential.

In this paper we introduce a set of best practices for high-stakes assessment in CBE, drawing from both the educational-measurement literature and current practices in prior-learning and CBE assessment. Broadly speaking, there are two areas in assessment design and implementation that require significant and sustained attention from test developers and program administrators: (1) validating the assessment instrument itself and (2) setting meaningful competency thresholds based on multiple sources of evidence. Both areas are critical for supporting the legitimacy and value of CBE credentials in the marketplace.

This paper therefore details how providers can work to validate their assessments and establish performance levels that map to real-world mastery, paying particular attention to the kinds of research and development common in other areas of assessment. We also provide illustrative examples of these concepts from prior-learning assessments (for example, Advanced Placement exams) and existing CBE programs. Our goal is to provide a resource to institutions currently developing CBE offerings and to other stakeholders—regulators and employers, for instance—who will encounter an increasing number of CBE programs.

Based on our review of the current landscape, we argue that CBE programs have dedicated most of their attention to defining discrete competencies and embedding those competencies in a broader framework associated with degree programs. Many programs clearly document not only the competencies but also the types of assessments they use to measure student proficiency. This is a good start.

We argue that, moving forward, CBE programs should focus on providing evidence that supports the validity of their assessments and their interpretation of assessment results. Specifically, program designers should work to clarify the links between the tasks students complete on an assessment and the competencies those tasks are designed to measure. Moreover, external-validity studies—relating performance on CBE assessments with performance in future courses or in the workplace—are crucial if CBE programs want employers to view their assessments and their competency thresholds as credible evidence of students' career readiness.

External validity is the central component of our recommendations:

1. CBE programs should clearly define their competencies and clearly link those competencies to material covered in their assessments.

2. To support valid test-score interpretations, CBE assessments should be empirically linked to external measures such as future outcomes.

3. Those empirical links should also be used in the standard-setting process so providers develop cut scores that truly differentiate masters from nonmasters.

4. In addition to rigorous test development and standard setting, CBE programs should continue to collect and monitor graduates' life outcomes in order to provide evidence that a CBE credential stands for a level of rigor and preparation equivalent to a traditional postsecondary degree.

Measuring Mastery: Best Practices for Assessment in Competency-Based Education

Katie Larsen McClarty and Matthew N. Gaertner

This paper is the third in a series examining competency-based higher education from a number of perspectives.

While college costs have risen dramatically over the past decade, degree completion rates have remained stubbornly flat, leading policymakers and advocates to look for new models of education that can reduce costs and raise productivity. Reformers have increasingly touted competency-based education (CBE) as a potential remedy for escalating prices and stagnant graduation rates.[1]

The case for CBE is intuitively appealing: Students can earn college credit by demonstrating competencies rather than accruing a certain amount of seat time, the conventional metric. In simple econometric terms, traditional higher education programs hold time constant (for example, students must complete 120 credit hours to earn a bachelor's degree) but allow the amount of demonstrated learning during that time to vary (for example, students can earn different course grades and still receive the same number of credit hours). CBE programs aim for the opposite: the standards for demonstrated learning are held constant, but the amount of time students must spend to reach them can vary.

CBE is particularly appealing for students whose work or family commitments make educational flexibility a priority. Such students represent a large and growing share of the college-going population. Twenty percent of undergraduate students work full time, with more than 70 percent working at least part time.[2] Nearly a quarter of undergraduate students are parents, and half of those are single parents.[3] Work and family priorities compete with class schedules and may make it difficult for some students to adhere to the seat time requirements of traditional education models where classes often meet in the middle of the day and in the middle of the week. CBE can help these students work at their own pace and on a more feasible schedule. And they can use the program to show they have mastered a predetermined set of competencies.

The idea of CBE is not new. In the 1970s, the US Department of Education Fund for the Improvement of Postsecondary Education made grants to support the development of new CBE programs at institutions that were already providing adult-learning programs. One grant recipient—a consortium of Minnesota community colleges—began developing a CBE program in 1973 and, two years later, 250 students across the St. Paul metropolitan area were enrolled. An evaluation of competency-based teacher education programs in Minnesota and Nebraska showed improved performance for beginning teachers, and higher levels of teacher and student satisfaction.[4]

Although CBE programs remained a small part of higher education for many years, their focus on student knowledge and outcomes rather than time spent in a traditional classroom led to advances in the movement to grant credit for prior learning. When Western Governors University (WGU) was founded in the late 1990s, it represented the first higher education institution to award degrees based solely on competencies. CBE programs are now firmly established elsewhere, at institutions such as Alverno College, Capella University, Excelsior College, Lipscomb University, and Southern New Hampshire University.

The emerging completion agenda has taken CBE from a niche market to the forefront of federal and state higher education policy discussions. In March 2013, the Department of Education announced that students

participating in approved CBE programs could be eligible for federal financial aid, echoing what advocates have been saying about the model for years:

> Competency-based approaches to education have the potential for assuring the quality and extent of learning, shortening the time to degree/certificate completion, developing stackable credentials that ease student transitions between school and work, and reducing the overall cost of education for both career-technical and degree programs. The Department plans to collaborate with both accrediting agencies and the higher education community to encourage the use of this innovative approach when appropriate, to identify the most promising practices in this arena, and to gather information to inform future policy regarding competency-based education.[5]

Students can currently receive federal financial aid under two types of CBE models. The first is a course-based model with credit equivalency. In this approach, student competencies are built into particular courses and then mapped back to credit hours. Although the credit hour is not the underlying metric of student learning, credit-hour equivalence is used to qualify students for financial aid. This was the original CBE model and is still the most popular. The second model, direct assessment, abandons consideration of credit hours altogether in favor of a direct measure of student learning such as projects, papers, examinations, presentations, performances, and portfolios. So far, though, regulators have only tentatively granted access to CBE models that are entirely divorced from the credit hour: only two institutions—Southern New Hampshire University and Capella University—have received both regional accreditor and Department of Education approval for direct-assessment programs.[6]

From a regulator's perspective, such caution is understandable given the pace of change and the calls for expansion.[7] Despite CBE's rising popularity, many important questions remain. A measure of learning is more intuitively appealing than a measure of time, but a CBE model is workable only insofar as its measures of learning yield trustworthy data about students' prospects for future success. Fortunately, CBE providers can help prove the value of the model by providing regulators and employers with clear, concrete evidence that their competencies and assessments truly differentiate students who have mastered necessary material from those who have not. Marshalling this evidence, in turn, requires the kind of best practices in assessment development, standard-setting processes, and evaluation that have been developed in psychometrics.

A CBE model is workable only insofar as its measures of learning yield trustworthy data about students' prospects for future success.

This paper therefore seeks to explore the current state of CBE assessment relative to best practices in assessment development and validation. We describe how prior-learning assessments have been implemented in higher education and how sound assessment principles and lessons learned have been or could be applied to CBE programs. We begin with a review of two frameworks: the first describes industry standards for developing and validating assessments, and the second focuses on determining mastery. Next, we apply each of the frameworks to existing prior-learning assessments and CBE programs, concluding with a set of recommendations for institutions implementing or planning to implement CBE programs.

The Common Elements of Competency-Based Education

CBE models can take a variety of forms, but most programs include two common elements: (1) a competency framework and (2) competency assessments. The competency framework describes the "skills, abilities, and knowledge needed to perform a specific task."[8] Competencies must be clearly defined, measurable, and related to the knowledge or skills needed for future endeavors, such as additional education or employment.[9] Often, competencies are specific to a particular course or degree program. For example, competencies in a public health

program may include being able to "identify public health laws, regulations, and policies related to prevention programs" or "use statistical software to analyze health-related data."[10] The second common element of CBE models is competency assessment. Because competency assessments are used to determine mastery and award credit, the value of CBE credentials hinges on the reliability and validity of those assessments.

Assessment quality has been an important research topic for as long as CBE programs have existed. In 1976, John Harris and Stephen Keller outlined several key considerations in competency assessment and concluded that "the major development effort in competency-based education should not lie in design of instructional materials but in design of appropriate performance assessments. Furthermore, institutions should not commit themselves to competency-based curricula unless they possess means to directly assess students' performance."[11]

Nearly 40 years later, that imperative persists. In Paul Gaston's book about higher education accreditation, he states: "Qualifying [CBE] programs should be expected to demonstrate that meaningful program-level outcomes are equivalent to those accomplished by more traditional means and, thereby, deserving of recognition through equivalent credentials."[12] The implications of this statement bear emphasis: Reliable assessment is a necessary but insufficient precondition for CBE program success. Programs must also produce students who are just as well prepared for future success as comparable students who earn credentials through more traditional avenues. It seems evident, then, that widespread acceptance and adoption of the CBE model will require high-quality competency assessments linked to meaningful labor market outcomes.

When developing competency assessments, there are two important stages. The first is assessment development and score validation—in other words, do scores on the assessment reflect the different levels of knowledge and skills that assessment designers are trying to measure? The second is determining how well a student must perform on the assessment in order to demonstrate competency—in other words, what is the cut score that separates the competent from the not-yet-competent? In this section we address each stage separately, drawing on best practices in each area.

Framework for Assessment Design. Assessment designers should start with the *Standards for Educational and Psychological Testing*, the book that describes industry standards for assessment development and validation.[13] The *Standards* provides guidance for developing and evaluating assessments and outline the types of evidence needed to support valid inferences about assessment results. Basically, an assessment is valid if there is appropriate evidence to support the intended score interpretations and the ways in which those who give the test will use it. Validity is obviously crucial for assessment development in CBE programs, where test scores may be used to confer not only course credits but also degrees or certificates.[14]

Imagine a test developed to measure a student's knowledge of public-health laws, regulations, and policies. Students with higher scores should exhibit a greater level of knowledge of public-health concepts. Their level of knowledge, as evidenced by their test scores, could be used to determine whether they are awarded competency credits in this area and, by extension, whether they are prepared for future endeavors in public health.

Although this understanding of test scores may seem intuitive, the ability to make valid inferences from assessment results relies on these simple axioms. For example, can the test developers demonstrate that knowledge of public-health laws, regulations, and policies—and not some irrelevant trait—explains test-score variability? Moreover, do higher scores relate to higher levels of subsequent job performance? Validation is the process of accumulating evidence to answer these fundamental questions. According to the *Standards*, validity evidence can come from five sources: (1) test content, (2) response processes, (3) internal test structure, (4) relations to other variables, and (5) test consequences.[15]

The first three sources of evidence generally reflect the test instrument itself, whereas the second two rely on data external to the assessment. Although not all sources of validity evidence may be present for every assessment, programs can make a strong validity argument by integrating evidence from multiple sources. For example, it is important to show that a competency-based assessment does test the knowledge and skills associated with the specified competency (evidence

based on test content). It is just as important, however, to show that students who score higher on the assessment also do well on other tasks, such as job performance, that require that competency (evidence based on relations to other variables). In a later section on assessment design in practice, we detail specific validity evidence that supports intended score interpretations for existing prior-learning and CBE assessments.

Framework for Determining Mastery. Once an assessment has been developed, test designers must establish cut scores to separate masters from nonmasters. In the case of CBE, the assessment cut scores distinguish those who receive credit (or various levels of credit) from those who do not. Because cut scores are central to the use and interpretation of CBE assessments, test designers must also gather validity evidence to support cut-score placement.

One particularly relevant approach for setting cut scores and determining mastery is Evidence Based Standard Setting (EBSS), which is especially useful when an assessment makes claims about future performance (for example, a test-taker's ability to pass future courses or succeed in the workplace).[16] In K–12 settings, EBSS approaches have been used to identify college-ready high school students by using data that link secondary school test scores with how those students perform once they reach college.[17] CBE credentials imply preparedness for future work, so EBSS may be similarly well suited to setting cut scores on CBE assessments.

In an EBSS approach, the judgments of subject-matter experts are combined with data from research studies to determine the cut scores for different performance levels. The five EBSS steps are described in detail in a subsequent section, but first, we turn to assessment design in practice.

Assessment Design in Practice

CBE assessment can take a variety of formats: objectively scored assessments (for example, those with multiple-choice or true-false questions), performance-based assessments (for example, those including essays, group projects, or simulated environments), and real-world observations (for example, preservice teachers in the classroom). Regardless of format, however, the credibility of inferences drawn from assessment results depends on evidence of their validity. A 2002 Department of Education review of CBE programs, for instance, stated that few programs report robust reliability or validity information. The authors note, "By attending to concerns about validity and reliability, institutions can glean meaningful information to improve their initiatives and to satisfy external demands for accountability."[18] In this section, we describe potential sources of such validity evidence and provide examples of evidence from CBE programs and prior-learning assessments. We also note examples of evidence that CBE providers could collect to validate and promote their model going forward.

Intended Score Interpretations and Test Use. The first step in developing an assessment and amassing the appropriate validity evidence is specifying the purpose of the assessment, or the intended interpretation of test scores for a given use. Once specified, that interpretation must be validated. Because CBE assessments provide evidence of student learning and are used to award credits, degrees, or other certifications—qualifications that students can take with them from one institution to another or from institution to employer—this evidence should theoretically be transportable across educational institutions and sectors. At present, the Department of Education notes that CBE programs in the United States are far from achieving this goal (although, to be fair, most traditional colleges suffer from the same portability challenges).[19]

To support portability, CBE programs should gather evidence corresponding to the five validity elements described in *Standards for Educational and Psychological Testing*. Specifically, CBE programs should

1. Clearly define the competencies;
2. Provide an explicit link between the skills measured by the assessments and those competencies;
3. Demonstrate that student behaviors or thought processes during testing reflect the competencies;

4. Relate performance on competency assessments with other measures of the same competencies; and

5. Document the empirical relationship between assessment scores and future outcomes (such as success in the workplace or attainment of a more advanced competency).

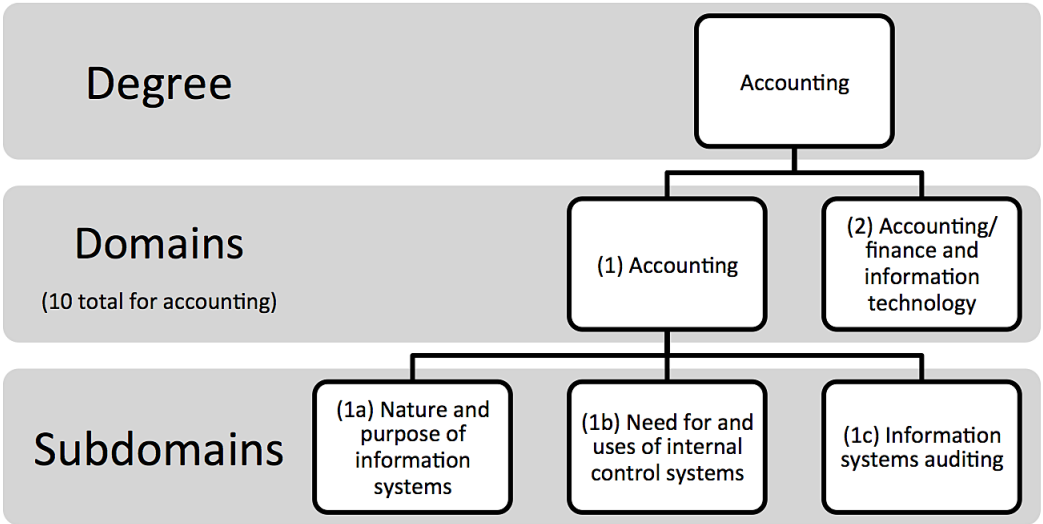
CBE programs must also provide detailed information about the intended interpretations and uses of their assessments. For example, Excelsior College's CBE nursing students are expected to demonstrate theoretical learning and clinical competence, including critical thinking, at a level required for beginning practice as an associate-degree-level registered nurse.[20] Accordingly, the Excelsior nursing assessments should be designed to measure students' theoretical knowledge, clinical competence, and critical thinking. To earn a CBE nursing degree, student performance on the assessments should be similar to that of nurses with associate degrees, and their performance on the job should be similar to that of other nurses at that level.

Defining Competencies. Perhaps the most important step in assessment design is defining the competencies. As Richard Voorhees has argued, competencies must be clearly defined and measurable; otherwise, they cannot be considered competencies.[21] Therefore, designers of CBE programs must clearly define the competency or set of competencies an assessment will measure. An exemplar in this area is WGU. For each degree WGU awards, a set of domains is specified. For example, a bachelor's of science in accounting consists of 10 domains: (1) accounting; (2) accounting/finance and information technology; (3) business law and ethics; (4) cost/managerial accounting; (5) economics, global business, and quantitative analysis; (6) foundations; (7) liberal arts; (8) marketing and communication; (9) organizational behavior and management; and (10) system administration and management.

For each domain, a set of subdomains elaborate the specific competencies that a student must demonstrate. The subdomains within accounting include "the student understands the nature and purpose of information systems; the student understands the need for and uses of internal control systems; and the

FIGURE 1

HIERARCHICAL STRUCTURE OF DEGREES, DOMAINS, AND COMPETENCIES



Source: The authors
Note: This structure is based on Western Governors University's bachelor's degree in accounting.

student understands information systems auditing." The domains and subdomains are jointly developed by subject-matter experts and employers in that field.[22] Figure 1 shows the hierarchical structure of the degree program, domains, and subdomains.

Alverno College has established four institution-wide learning outcomes, which it has regularly expanded and revised, and today the institution lists eight competencies required of all students.

A second example of degree-level competency specification comes from Alverno College, which began exploring CBE programs in the late 1960s. First, educators used a faculty survey to capture the learning outcomes that professors saw as critical for individual courses and academic departments. A subsequent analysis identified the similarities across courses and departments, from which Alverno established four institution-wide learning outcomes. Alverno College regularly expanded and revised those learning outcomes, and today the institution lists eight competencies required of all students: communication, analysis, problem solving, value in decision making, social interaction, developing a global perspective, effective citizenship, and aesthetic engagement.[23]

Another approach to defining competencies in CBE programs is to establish a large master set of competencies and then require students to demonstrate proficiency in a subset depending on the degree program or job requirements. Lipscomb University uses this approach. Lipscomb licensed the Polaris business competency framework, which defines 41 competencies across seven general categories: interpersonal, communication, management, leadership, conceptual, personal, and contextual.[24]

The Polaris system has been implemented by numerous companies across a variety of industries to hire staff, provide training, and develop leaders. When Lipscomb applied the Polaris competency model, feedback from local business and industry stakeholders suggested that 17 of the 41 competencies would be relevant and appropriate qualifiers for an undergraduate degree. Lipscomb now requires students seeking a CBE undergraduate degree to demonstrate mastery of these 17 competencies.

The Validity of the Test Instrument. Before educators can use an assessment to award credit or degrees for demonstrated competencies, they must determine that the assessment is a valid measure of those competencies. That means that (1) the test must fully measure the competency, (2) the processes students use to complete the assessment tasks must be an authentic reflection of the competency, and (3) students would receive the same test results if they were to take a different form of the test scored by different raters. These ideas correspond with validity evidence based on test content, response processes, and internal structure, respectively. Each will be briefly described in the following paragraphs.

Evidence Based on Test Content. Once program designers have clearly defined relevant competencies, they should collect evidence that their test content fully reflects those competencies. Specifically, providing validity evidence based on test content means showing the relationships between test questions or tasks and the defined competencies. Test developers must consider how well the breadth and depth of their test relates to defined competencies.

The Advanced Placement (AP) program is one example. Its test content validity evidence is grounded in a process known as evidence-centered design (ECD).[25] Using ECD, assessment designers connect three components: (1) the intended claims about students' content knowledge and skill proficiency, (2) the observable evidence a student would need to produce to support those claims, and (3) the tasks or situations that would elicit that evidence. By designing assessment tasks that enable students to demonstrate the relevant knowledge and skills, test developers provide evidence of validity.

An example from the AP Chemistry assessment starts with the following claim: "Students can apply mathematics in which they evaluate the reasonableness of quantities found in stoichiometric calculations."[26]

Examples of supporting evidence would include the correctness of a chemical equation, chemical formulas, application of mathematical routine, or coefficients interpreted as mole ratios. Assessment items can then be written such that students demonstrate the application of a correct chemical formula or interpret the coefficients of a problem as mole ratios. Test content is thereby directly linked to defined competencies through the ECD process.

Although not necessarily developed with ECD principles in mind, some CBE assessments make an explicit link between test content and defined competencies. For example, in Southern New Hampshire University's direct-assessment CBE program, students show proficiency in various competencies through authentic project tasks. Students are able to select from multiple simple projects that assess one competency at a time or a single complex project that assesses multiple competencies. A simple project assessing students' ability to write a paragraph involves describing a recent purchase—specifically, why the item was purchased and why it was selected over other items.

A more complex project assessing multiple competencies, on the other hand, requires a student to write a formal memo to his or her boss evaluating two vending machine companies and recommending one over the other. The vending machine recommendation project assesses five competencies: (1) can use logic, reasoning, and analysis to address a problem; (2) can write a business memo; (3) can use a spreadsheet to perform calculations; (4) can synthesize material from multiple sources; and (5) can evaluate information and its sources critically.[27] This explicit project-to-competency link provides strong validity evidence based on test content for Southern New Hampshire's CBE program.

Evidence Based on Response Processes. Students' response processes—that is, the thoughts, behaviors, and actions required of a student to complete an assessment—are another source of validity evidence, usually gathered during initial test development. For example, students taking the AP Music Theory course and assessment must demonstrate a variety of skills through different processes. Aural skills are measured through exercises requiring students to listen to a piece of music and write the notation on a staff.[28] Sight-singing skills, on the other hand, are best measured through a performance assessment where the student sings a set of notes into a recorder. Using novel pieces of music and sets of notes helps ensure that the assessments are measuring specifically aural skills or sight singing rather than memory or musical experience.

CBE programs can and should gather similar evidence. For example, in Excelsior College's nursing program, a computer-based exam is given to assess nursing theory, but critical thinking and clinical reasoning are measured through simulated clinical experiences and actual performance in a clinical setting.[29] While it seems preferable to assess clinical reasoning in a clinical setting, assessment designers must clearly describe how adequate reasoning skills are demonstrated (or insufficient reasoning skills identified) in such a test-taking scenario. In the case of this nursing exam, establishing explicit links between the desired thinking and reasoning processes and successful task completion would provide validity evidence based on response process.

Evidence Based on Internal Structure. A third type of validity evidence is based on the internal structure of the assessment—that is, how the assessment items or tasks relate to each other. For example, the developers of the AP World Languages and Cultures assessments hypothesized that their tests measured four factors: reading, writing, listening, and speaking. Factor analysis (a common and useful tool for determining the number of factors a test measures) supported their hypothesis.[30]

Another way to consider this is to compare test structure across different examinee groups. For example, College Board conducted several studies to determine whether the AP World Language and Cultures exams kept their four-factor structure for native speakers, bilingual students, and second-language learners. Results supported a similar factor structure across all population groups.[31]

The most common way to demonstrate validity evidence based on internal structure is through reliability. There are different ways to measure different types of reliability, including test-retest (where students take the same test form on different occasions), internal consistency (which measures the extent to which students respond similarly to items within a single test

form), and inter-rater reliability (where two or more raters evaluate the same student performance on a test). Students should receive approximately the same score if they take a test multiple times, regardless of the test form administered or the raters scoring it. Using Cronbach's alpha (a reliability statistic that ranges from 0 to 1.0) as a measure of internal consistency, for example, values above 0.80 are considered acceptable, although most standardized tests typically have values above 0.90.[32] The AP program reports the reliability of each section (multiple choice and free response), of the raters, of the composite score, and of the subsequent score classification.[33]

All three of these analyses could be applied to CBE programs. Program designers could apply statistical analyses to the WGU assessments (described previously in this section) to determine whether their structure reflects the domain and subdomain structure specified in the competency frameworks. Additional analyses could help them evaluate whether the structure is consistent across relevant population groups. Finally, CBE program designers should always report reliability statistics when tests are used for high-stakes purposes.

Although many CBE programs report developing reliable and valid assessments, reliability statistics are rarely publicly documented. Some programs rely on assessments developed by external organizations, and those organizations typically provide reliability information for their instruments. For example, the reliability of the Polaris assessments (used by Lipscomb University) exceeds the 0.80 threshold for all but a few dimensions. Many more institutions, however, are developing their own assessments and should work to provide their own reliability evidence.

Validity Associated with External Evidence. While we just outlined three sources of validity related to the test itself—test content, response processes, and internal structure—this section describes sources based on external evidence. External-validity evidence is critical to supporting the claims that CBE programs can make about the relationship between their measures of competence and workplace success, and about comparability of graduates from CBE and non-CBE programs.

Concurrent Validity Evidence. Validity evidence based on relationships with other variables can come at two points. First, educators could compare assessment results with other measures collected concurrently. For example, students completing a college algebra course may be administered a College-Level Examination Program (CLEP) exam to evaluate the relationship between performance on CLEP and performance in the course. A strong positive relationship between test performance and course performance would support using the CLEP test to place out of college algebra. Indeed, there is evidence that CLEP scores are moderately correlated with college course grades.[34]

External-validity evidence is critical to supporting the claims that CBE programs can make about the relationship between their measures of competence and workplace success, and about comparability of graduates from CBE and non-CBE programs.

Similar evidence is limited for CBE programs. Examples do exist, however, from programs outside the United States. The National Cancer Action Team in England developed a competency assessment tool for technical surgical performance. Test developers validated their assessment tool against other measures of examinee performance—namely, a measure of observed errors. As one would hope, scores on the competency assessment tool were inversely related to the number of surgical errors.[35]

Predictive Validity Evidence. In addition to concurrent validity evidence, predictive validity evidence is critical when assessment scores will be used to predict a future outcome. Performance on AP exams, for example, should predict postsecondary outcomes. Accordingly, College Board has provided evidence that, after controlling for SAT scores and high school grades, students who scored higher on the AP exam had higher levels

of college success (in other words, higher grades, retention, and selectivity) than lower-scoring AP students or students not taking the AP exam.[36]

Colleges also use CLEP exams to award credit and therefore require similar validity evidence. That body of evidence suggests that students who receive college credit via CLEP perform comparatively well in their subsequent college courses: CLEP students typically have higher grade-point averages (GPAs) and earn a higher proportion of As and Bs relative to their classmates.[37] Even after controlling for prior achievement and demographic characteristics, CLEP students had higher GPAs than non-CLEP students and graduated sooner, enrolled in fewer semesters, and graduated with fewer credits.[38]

Hopefully as years pass and CBE programs mature, more institutions undertake and publish rigorous validity studies to establish a research base commensurate with CBE's growing popularity.

These studies provide strong evidence of validity based on test consequences. Similar performance patterns in subsequent courses helps demonstrate that students who succeed on a placement exam have indeed mastered the requisite skills; this is the evidentiary *sine qua non* for prior-learning assessments. For CBE programs to become widely accepted as an alternative path for earning a college degree, the programs must likewise provide evidence that they are just as good as corresponding traditional degree programs at imparting—or at least measuring—the relevant knowledge and skills.

Although such external-validity data for CBE assessments is relatively scant, some programs are developing infrastructure to support these important analyses. Lipscomb University students, for example, are rated by their employers at the beginning and end of the CBE program. Employers' ratings at the beginning of the CBE program could provide concurrent evidence when linked with students' initial performance on CBE assessments. Further, employers' postprogram ratings could provide evidence of the CBE assessments' predictive value.

Other, more mature CBE programs do report limited information related to later-life outcomes. For example, on its website WGU reports that its senior students performed better than students at 78 percent of institutions participating in the Collegiate Learning Assessment, a standardized measure of critical thinking and communication.[39] In addition, 94 percent of employers felt that WGU graduates performed at least as well as graduates from other institutions. In fact, 53 percent of employers reported higher performance from the WGU graduates.[40]

Excelsior College also reports outcomes data in terms of standardized-test performance and subsequent job performance. Graduates from the Excelsior nursing program pass the nursing licensure exam at rates comparable to the national average. Once employed, 82 percent of surveyed nurse supervisors rated Excelsior nursing graduates similar or higher in terms of clinical competency compared to other associate-degree-level nursing graduates.[41]

Posting student outcomes to a website or publishing job performance results via commissioned reports is a step in the right direction. But the educational research community needs more examples similar to those provided by Excelsior College and WGU. Furthermore, submitting claims about student outcomes to rigorous scientific peer review could substantially expand the CBE knowledge base and allow policymakers to fairly assess the value these programs provide. While that kind of research takes time, we hope that as years pass and CBE programs mature, more institutions undertake and publish rigorous validity studies to establish a research base commensurate with CBE's growing popularity.

Determining Mastery

The previous section focused on assessment design and the need to collect validity evidence for assessment results. In this section we focus on the equally important task of determining the level of performance

required to receive credit. In educational assessment, this is known as standard setting, a process that is common in primary and secondary schooling, but less frequently discussed in higher education. Standard setting is the process of defining discrete levels of achievement on an assessment and setting cut scores to separate those levels. In some cases, such as with licensure exams, two performance levels are sufficient: pass or fail. In other cases, more levels are useful to further differentiate performance. AP exams, for example, have five score points.

Standard setting not only relates assessment scores to performance levels, but also determines which performance levels are sufficient to receive credit. As described earlier, EBSS may provide a particularly attractive standard-setting framework for both prior-learning and CBE programs, because each performance level in those programs is associated with not only a competency level but also an empirical likelihood of future success. The following paragraphs describe each of the five steps of EBSS:

1. Define the relevant outcomes;

2. Design appropriate studies;

3. Conduct studies and synthesize results;

4. Stakeholder review and recommendations; and

5. Ongoing monitoring.[42]

Step 1: Define the Relevant Outcomes. The first step in EBSS is defining the competencies and corresponding indicators of future success for each performance level. For example, students generally need to demonstrate specific knowledge and skills to earn an AP exam score of 4—a claim about competency. Moreover, students who score a 4 would typically earn an A-, B+, or B in a corresponding college course—a claim about future success.

CBE program designers have established similar definitions. Lipscomb University's CBE program, for example, has four levels for each competency: basic/elementary, proficient practitioner, exceptional/expert, and master/guru. Each performance level is associated with a particular set of knowledge, skills, and behaviors. Lipscomb's "influence" competency, required for an undergraduate CBE degree, provides a good example. Students at the basic/elementary level of the influence competency are responsive: they acknowledge requests quickly, listen attentively, and gain respect and admiration. At the next level, proficient practitioners are reliable team leaders who identify and communicate compelling motivators. They adjust their influence style to meet the needs of individual team members and offer recognition and encouragement to keep the team moving forward.

The exceptional/expert influencer communicates a legitimate, consistent agenda across a variety of functions, understands power dynamics and the responsibilities of leadership, clearly articulates situational advantages, and validates potential concerns. Finally, individuals at the master/guru level develop and implement appropriate and creative recognition, rewards, and incentives to activate an organization. They influence all levels of the organization and external stakeholders through strong communication, impactful messages, and personal appeal. In addition, masters/gurus remain persistently optimistic, particularly in the face of challenges.[43]

Lipscomb's CBE program has also made associations with external measures. While each of these four categories describes a distinct level of competence, each is also linked to success in various tiers of employment. For example, students at the basic/elementary level are ready to become entry-level, individual contributors, while proficient practitioners are prepared for supervisor or entry-level manager positions. The exceptional/expert-level competencies are needed for functional managers or managers of managers. Finally, strategic leaders or corporate executives are associated with the master/guru level of performance.[44]

CBE program designers must also consider how many distinct performance categories can be clearly differentiated by their assessments, and the consequences of landing in any given level. Many CBE programs divide their assessments' scales into two levels (one in which students receive credit, and one in which they do not), but such a stark dichotomy is not required. Instead, different performance levels could translate to different numbers of credits awarded, or CBE programs

could establish graduated distinctions for exceptional performance (similar to course grades in traditional degree programs).

Step 2: Design Appropriate Studies. In step two, CBE program leaders must develop research studies that can produce evidence for the claims implied by their performance levels. For AP exams, college professors and high school AP course teachers first described what students should be able to do at each level and then estimated how many test questions a student would need to correctly answer to attain each score point. But to support claims about future outcomes, leaders had to design research studies that compared college course performance with AP exam scores.[45]

To support the competency claims made by CBE programs, assessment designers have implemented similar processes. In WGU's nursing program, for instance, a panel of university faculty and external experts reviewed the objectively scored assessment items and indicated how they thought a student with sufficient mastery of the competency would perform.[46] Similarly, designers of the Lipscomb University Adult Degree Program gathered recommendations from several different groups, including faculty and local employers, who recommended competency levels appropriate for an undergraduate degree, based on their expert knowledge of course and job requirements.

To our knowledge, however, neither university designed studies to support the claims about future outcomes as part of its initial standard-setting process. To be sure, this is typical of most standard-setting processes for educational assessments. Initial standards are often set based on expert judgments, while designers collect validity data based on relationships with future outcomes or testing consequences after the fact.

However, when assessments are linked to significant claims about future outcomes—as is the case for both prior-learning assessments and CBE programs—we argue that, where possible, program designers should seek out empirical evidence relating assessment results to external outcomes to inform the initial standard setting. Establishing these external links a priori takes time and careful planning but has proven feasible in large- and small-scale testing programs.[47]

Step 3: Conduct Studies and Synthesize Results. In step three, program designers carry out the research studies designed in step two and then combine and synthesize the results. For each AP exam, for instance, the College Board conducts a college comparability study, administering a shorter version of the AP exam to students enrolled in corresponding introductory college courses.[48] Students' performance on the AP exam is then compared to their course grades.

To establish empirical links between test scores and relevant outcomes, AP scores are averaged within discrete college grades. (For example, what is the average AP score for students earning an A-, B+, or B in the course?) Average performance can support claims about future outcomes at each of the five AP performance levels. In general, an AP score of 5 is equivalent to a college course grade of A; a 4 maps to college grades of A–, B+, and B; and a 3 maps to B–, C+, and C.[49]

This kind of external-validity research is rare in CBE programs. Few programs have linked performance on their assessments with future outcomes, so these links are also absent during the standard-setting process. This is a clear area for improvement: when CBE programs set minimum test scores for course credit, external data linking those judgments to future performance should play a central role.

Step 4: Stakeholder Review and Recommendations. In step four, stakeholders review both the assessments and the study evidence to determine the score that best differentiates students who have demonstrated mastery from those who have not (or students in one performance level from those in an adjacent level). For the AP program, the most relevant stakeholders are high school AP teachers and college faculty. For CBE programs, relevant stakeholders include not only the faculty who will implement the programs and their assessments but also colleagues, employers, or industry representatives who will be hiring CBE graduates.

For the AP program, stakeholders can consider information both from experts' judgments on the assessment items and from student performance in college courses. There are general guidelines about the relationship between AP exam performance and college course grades, but the stakeholders must also consider

variation among college courses and the judgments made by the expert reviewers in recommending cut scores for each performance level.[50]

In addition, stakeholders can use college comparability study results to consider the consequences of different cut-score placements. Because most colleges use an AP score of 3 or higher to award college credit, this cut score is particularly important. If the cut score is set too high, many students would still be required to complete an entry-level course even though they already have the knowledge and skills necessary to perform well in that course. If the cut score is set too low, students may place out of the entry-level course but not have the knowledge and skills to be successful in a subsequent course. The goal is finding the Goldilocks cut score that is not so high that overprepared students are unfairly sent to introductory courses, but not so low that underprepared students are allowed to skip material they have not yet mastered.

Stakeholder review is also an important step in setting standards for CBE programs. Stakeholders usually have a voice in defining the original competency framework; they should also be involved in determining the level of competency required for a credential. Furthermore, when setting standards for K–12 assessments, test developers typically publish a technical report describing the composition of stakeholders and their role in the standard-setting process. Technical reports are common practice in assessment development. Reports provide transparency and allow external review of test development processes and the associated validity evidence. CBE providers should make efforts to publish similar technical documentation for their assessments.

Step 5: Ongoing Monitoring. CBE program designers should collect data throughout the life of an assessment program to provide continuing support for the interpretation of scores and performance-level classifications. As appropriate, assessment developers and stakeholders may revisit and revise performance-level cut scores to reflect new evidence generated from multiple student cohorts progressing through a course or academic unit. For AP, this means additional iterations of the studies noted previously, where entry-level college course grades are linked to AP scores. Obviously, higher AP scores should continue to predict better grades in entry-level courses. Furthermore, if students receive AP credit and skip an entry-level course, in subsequent courses they should perform as well as comparable students who did not advance via AP credit.

Meanwhile, programs that do not include outcomes-based study evidence in the initial standard-setting process should monitor (and, if necessary, revise) cut scores after they have collected and analyzed such data. For example, Lipscomb can use employer ratings once graduates return to the workforce to evaluate whether reaching the proficient practitioner level predicts successful job performance, or whether requiring exceptional/expert-level performance might be more prudent for more demanding jobs. Likewise, Excelsior College can monitor the relationship between student performance on CBE nursing assessments and subsequent performance on the nursing licensure exam to ensure that students who pass the CBE program assessments are at least as well prepared for the licensure exam as students who opted for a traditional program. As CBE programs continue to mature, they should continue to gather data from students and graduates for use in regular reviews of their competency thresholds.

Conclusion and Recommendations

Competency-based education programs may well represent a viable alternative pathway to a postsecondary degree. Ideally, CBE students would progress at their own pace and demonstrate mastery of important competencies, free from the restrictions of traditional seat-time requirements. This would allow graduates to clearly describe (and provide evidence for) the knowledge, skills, and abilities they demonstrated to earn their degree. Employers could match their needs to candidates with relevant competencies. Practically speaking, though, the credibility of CBE credentials in the marketplace (and therefore the viability of the CBE model in general) rests on reliable and valid assessments with evidence-based performance levels.

Based on our review, it seems that most CBE efforts to date have focused on defining the competencies and developing the competency frameworks associated with various degree programs. Many programs have clear documentation of the competencies they seek to

teach and measure and the types of assessments they will use to determine mastery. Their next step should be to provide more specific documentation linking assessment tasks (such as test questions) with the competencies those tasks are designed to measure.

More importantly, CBE programs would be wise to begin longitudinal research linking their assessments to other relevant student outcomes, such as job performance. This type of evidence is crucial for establishing the validity of both CBE assessments and the cut scores separating those who receive credit from those who do not.

CBE programs would be wise to begin longitudinal research linking their assessments to other relevant student outcomes, such as job performance.

Throughout this paper, we have described the assessment industry's evidence standards and the current state of CBE assessment relative to those basic principles. By way of summarizing our observations and prodding further research, we conclude with some recommendations for institutions that currently offer (or are considering developing) CBE programs:

1. **Clearly define competencies, and document evidence that assessments fully measure those competencies**. Contemporary higher education debates are focused on the knowledge, skills, and abilities that college graduates should possess. The Lumina Foundation's Degree Qualifications Profile and the Association of American Colleges & Universities' Essential Learning Outcomes have set out to identify competencies for general education. So have several of the programs mentioned in this paper. Those competencies, however, must be defined with enough detail that they can be measured. Although many current CBE programs have detailed descriptions of the competencies, there is less documentation linking those competencies to the assessments that measure them. Such clarity would not only provide validity evidence but would also improve transparency around the processes and expectations of CBE programs.

2. **Conduct research to relate CBE assessments to other assessments measuring similar competencies and to future outcomes that assessments are designed to predict**. Many CBE programs develop their own assessments and have good reasons for not wanting to adopt large-scale standardized tests. But it is no less important that these local assessments be validated against other measures. CBE programs could collaborate to collect the necessary evidence. For example, students could complete both Alverno College's problem-solving measures and Lipscomb University's problem-solving and decision-making competency assessments. Collaborating institutions should not expect scores to be identical, but where competencies are conceptually related, the assessments of those competencies should show an empirical relationship.

3. **Use the results of empirical research studies in the initial standard-setting process**. Data relating CBE assessment scores to other outcomes should be used not only to validate the assessment post hoc but also to set competency standards a priori. Although performance standards must usually be set before longitudinal data linking assessment performance with future outcomes can be collected, there are viable alternatives for gathering outcomes data. CBE assessments could be administered to employees currently working in fields relevant for the assessment. For example, WGU offers degree programs in education, business, information technology, and health care. Entry-level workers in those fields could complete WGU's CBE assessments, and those students' on-the-job performance could be compared to their performance on CBE assessments. These empirical links could be evaluated in conjunction with the expert judgments currently collected, helping bolster the validity evidence supporting chosen performance levels.

4. **Continue to gather and report validity evidence for CBE assessments and performance standards, including comparisons of student outcomes against relevant comparison groups**. For CBE programs to be viewed as an attractive alternative to traditional programs, students and employers need evidence that (1) CBE graduates possess the same knowledge and skills as comparable traditional graduates and (2) CBE graduates are equally successful after graduation. These outcomes could be measured in terms of subsequent academic performance or through job attainment, job performance, occupational prestige, or earnings. Some CBE programs may be collecting these data already; they should focus on rigorous analysis and publication. Other programs will need to develop the necessary infrastructure and timelines to start data collection. It will take time to gather robust long-term outcomes data, but these data can provide compelling evidence for the effectiveness of CBE programs and support their continued growth.

Notes

1. US Department of Education, "Applying for Title IV Eligibility for Direct Assessment (Competency-Based) Programs," March 19, 2013, http://ifap.ed.gov/dpcletters/GEN1310.html.

2. Jessica Davis, "School Enrollment and Work Status: 2011," US Census Bureau, October 2012, www.census.gov/prod/2013pubs/acsbr11-14.pdf.

3. Center for Law and Social Policy, "Yesterday's Nontraditional Student Is Today's Traditional Student [Fact Sheet]," June 2011, www.clasp.org/resources-and-publications/publication-1/Nontraditional-Students-Facts-2011.pdf.

4. Roland L. Peterson, "Review and Synthesis of Research in Vocational Teacher Education," Ohio State University, Center for Vocational Education, 1973, http://files.eric.ed.gov/fulltext/ED087898.pdf.

5. White House, Office of the Press Secretary, "Fact Sheet on the President's Plan to Make College More Affordable: A Better Bargain for the Middle Class," August 22, 2013, www.whitehouse.gov/the-press-office/2013/08/22/fact-sheet-president-s-plan-make-college-more-affordable-better-bargain-.

6. Patricia Book, "All Hands on Deck: Ten Lessons from Early Adopters of Competency-Based Education," Western Interstate Commission for Higher Education, May 2014, http://wcet.wiche.edu/wcet/docs/summit/AllHandsOnDeck-Final.pdf.

7. In 2014, the Department of Education's inspector general issued a report warning of the potential for waste, fraud, and abuse to result from the department's approval of direct assessment programs. See US Department of Education, Office of the Inspector General, "Final Audit Report," September 30, 2014, www2.ed.gov/about/offices/list/oig/auditreports/fy2014/a05n0004.pdf.

8. Ibid, 7.

9. Sally Johnstone and Louis Soares, "Principles for Development Competency-Based Education Programs," *Change: The Magazine of Higher Learning* 46, no. 2 (2014): 12–19.

10. Council on Education for Public Health, "Competencies and Learning Objectives," June 2011, http://ceph.org/assets/Competencies_TA.pdf.

11. John Harris and Stephen Keller, "Assessment Measures Needed for Competency-Based Higher Education," *Peabody Journal of Education* 53, no. 4 (1976): 241–47.

12. Paul Gaston, *Higher Education Accreditation: How It's Changing, Why It Must* (Sterling, VA: Stylus Publishing, 2014).

13. *Standards for Educational and Psychological Testing* (Washington, DC: American Educational Research Association, 2014).

14. Ibid. According to Standard 1.0 of the *Standards for Educational and Psychological Testing*, "Clear articulation of each intended test score interpretation for a specified use should be set forth, and appropriate validity evidence in support of each intended interpretation should be provided."

15. Ibid.

16. Katie Larsen McClarty et al., "Evidence-Based Standard Setting: Establishing a Validity Framework for Cut Scores," *Educational Researcher* 42, no. 2 (2013): 78–88.

17. Edward Haertel, Jennifer N. Beimers, and Julie A. Miles, "The Briefing Book Method," in *Setting Performance Standards: Foundations, Methods, and Innovations*, ed. Gregory Cizek (New York, NY: Routledge, Second Edition, 2012); and Kimberly O'Malley, Leslie Keng, and Julie A. Miles, "From Z to A: Using Validity Evidence to Set Performance Standards," in *Setting Performance Standards: Foundations, Methods, and Innovations*, ed. Gregory Cizek (New York, NY: Routledge, Second Edition, 2012).

18. Elizabeth A. Jones, Richard A. Voorhees, and Karen Paulson, "Defining and Assessing Learning: Exploring Competency-Based Initiative," Institute of Education Sciences, National Center for Education Statistics, November 6, 2002, http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2002159.

19. Ibid.

20. Rebecca Klein-Collins, "Competency-Based Degree Programs in the US: Postsecondary Credentials for Measurable Student Learning and Performance," Council for Adult and Experiential Learning, 2012, www.cael.org/pdfs/2012_competencybasedprograms.

21. Richard A. Voorhees, "Competency-Based Learning Models: A Necessary Future," *New Directions for Institutional Research* 110 (2001): 5–13.

22. Klein-Collins, "Competency-Based Degree Programs."

23. Ibid.

24. Council for Adult and Experiential Learning, "Customized, Outcome-Based, Relevant Evaluation (CORE) at Lipscomb University: A Competency-Based Education Case Study," 2014, www.cael.org/cael_lipscomb_case_study.

25. Robert J. Mislevy and Geneva D. Haertel, "Implications for Evidence-Centered Design for Educational Assessment," *Educational Measurement: Issues and Practice* 25, no. 4 (2006): 6–20.

26. Maureen Ewing et al., "Representing Targets of Measurement within Evidence-Centered Design," *Applied Measurement in Education* 23, no. 4 (2010): 325–41.

27. Jennifer Share, "College for America: A New Approach for a New Workforce That Is Accessible, Affordable, and Relevant," in *2013 CAEL Forum & News: Competency-Based Education,* ed. Diana Bamford-Rees et al. (Council for Adult and Experiential Learning, 2013), www.cael.org/pdfs/cael_competency_based_education_2013.

28. College Board, "Music Theory Course Description," 2012, https://secure-media.collegeboard.org/ap-student/course/ap-music-theory-2012-course-exam-description.pdf.

29. Klein-Collins, "Competency-Based Degree Programs."

30. April Ginther and Joseph Stevens, "Language Background, Ethnicity, and the Internal Construct Validity of the Advanced Placement Spanish Language Examination," in *Validation in Language Assessment,* ed. Antony John Kunnan (Mahwah, NJ: Lawrence Erlbaum, 1998); and Rick Morgan and John Mazzeo, "A Comparison of the Structural Relationships among Reading, Listening, Writing, and Speaking Components of the AP French Language Examination for AP Candidates and College Students," Educational Testing Service, 1988.

31. Ibid.

32. Robert F. DeVellis, *Scale Development: Theory and Applications* (Newbury Park, NJ: Sage Publications, 1991).

33. Brent Bridgeman, Rick Morgan, and Ming-mei Wang, "Reliability of Advanced Placement Examinations," Educational Testing Service, 1996.

34. Amiel T. Sharon, "The Use and Validity of the GED and CLEP Examinations in Higher Education" (presentation, American Personnel and Guidance Association Annual Convention, Atlantic City, NJ, April 1971).

35. Danilo Miskovic et al., "Is Competency assessment at the Specialist Level Achievable? A Study for the National Training Programme in Laparoscopic Colorectal Surgery in England," *Annals of Surgery* 257 (2013); 476–82.

36. Krista Mattern, Emily Shaw, and Xinhui Xiong, "The Relationship between AP Exam Performance and College Outcomes," College Board, 2009, https://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2009-4-relationship-between-ap-exam-performance-college-outcomes.pdf.

37. Brad Moulder, Abdulbaset Abdulla, and Deanna L. Morgan, "Validity and Fairness of CLEP Exams," College Board, 2005, http://media.collegeboard.com/digitalServices/pdf/clep/validity-fairness-clep-exam.pdf.

38. Carol Barry, "A Comparison of CLEP and Non-CLEP Students with Respect to Postsecondary Outcomes," College Board, 2013, http://media.collegeboard.com/digitalServices/pdf/clep/clep_research_report.pdf.

39. Western Governors University, "WGU Student and Graduate Success," http://texas.wgu.edu/about_wgu_texas/learning_results.

40. Western Governors University, "WGU Is Focused on Student and Graduate Success," www.wgu.edu/about_WGU/graduate_success.

41. Li Gwatkin, Mary P. Hancock, and Harold A. Javitz, "As Well Prepared, and Often Better: Surveying the Work Performance of Excelsior College Associates Degree in Nursing Graduates," SRI International, November, 25, 2009, https://my.excelsior.edu/documents/78666/102207/Work_Performance_of_Excelsior_Associate_Nursing_Graduates.pdf/357ed375-41ce-436a-be9e-73a96a34ec51.

42. McClarty et al., "Evidence-Based Standard Setting."

43. Organization Systems International, "OSI Polaris® Competency Continuums."

44. Other CBE programs have developed similarly detailed definitions of performance levels. For the sake of parsimony,

we will not describe each in detail here. For a thorough description of the performance levels for Alverno College's problem-solving competency and Tusculum College's coherence competency, see Rebecca Klein-Collins, "Competency-Based Degree Programs."

45. For an example of this type of study, see Brian F. Patterson and Maureen Ewing, "Validating the Use of AP exam Scores for College Course Placement," College Board, 2013, http://research.collegeboard.org/sites/default/files/publications/2013/7/researchreport-2013-2-validating-AP-exam-scores-college-course-placement.pdf.

46. Jan Jones-Schenk, "Nursing Education at Western Governors University: A Modern, Disruptive Approach," *Journal of Professional Nursing* 30, no 2. (2014): 168–74.

47. For example, a consortium of states developing an Algebra II assessment as part of the American Diploma Project used this approach, and Texas also used this approach in developing standards for its statewide assessment program. See Haertel, Beimers, and Miles, "The Briefing Book Method"; and Texas Education Agency, "STARR Performance Standards," http://tea.texas.gov/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=25769804117&libID=25769804117.

48. Patterson and Ewing, "Validating the Use of AP Exam Scores for College Course Placement."

49. College Board, "Music Theory Course Description."

50. Deborah Lokai Bischof et al., "Validating AP Modern Foreign Language Exams through College Comparability Studies," *Foreign Language Annals 37,* no. 4 (2004): 616–22.

Other Papers in This Series

- **The Landscape of Competency-Based Education: Enrollments, Demographics, and Affordability**, *Robert Kelchen*
- **Employer Perspectives on Competency-Based Education,** *Chip Franklin and Robert Lytle*

About the Authors

Katie McClarty, director of the Center for College & Career Success in Pearson's Research & Innovation Network, leads a team of researchers who plan and execute research in support of the center's mission, which is to identify and measure the skills needed to be successful in college and careers, determine pathways for students to be college and career ready, track their progress along those pathways, and evaluate effective ways to keep students on track. McClarty has authored papers and presentations related to college readiness, standard setting, assessment design, and talent development. Her work has been published in journals such as the *American Psychologist, Research in Higher Education, Educational Measurement: Issues and Practice,* and *Educational Researcher.*

Matthew Gaertner is a senior research scientist at the Center for College & Career Success in Pearson's Research & Innovation Network. His methodological interests include multilevel models, categorical data analysis, and item response theory. Substantively, his research focuses on the effects of educational policies and reforms on disadvantaged students' access, persistence, and achievement. Gaertner's work has been published in *Harvard Law Review, Harvard Educational Review, Educational Evaluation and Policy Analysis, Research in Higher Education*, and *Educational Measurement: Issues and Practice*. He was awarded a Spencer Foundation Dissertation Fellowship and an Association for Institutional Research Dissertation Grant. He also received the 2013 and 2011 Charles F. Elton Best Paper Awards from the Association for Institutional Research.

Acknowledgments

We would like to express our gratitude to Charla Long at Lipscomb University for providing detailed information about the school's CBE program. We also thank Andrew Kelly and Rooney Columbus of AEI for supporting this project and providing feedback on earlier drafts. All remaining errors are our own.