Articles

Assessing Developmental Assessment in Community Colleges

Community College Review
39(4) 327–351

DOI: 10.1177/0091552111426898
http://crw.sagepub.com
SAGE

Katherine L. Hughes[1] and Judith Scott- Clayton[1]

Abstract

For many students entering community colleges, the first stop on campus is to an assessment center. More than half of these students will be placed into developmental education as a result of their scores on reading, writing, and mathematics entry assessments, yet there is little evidence that this improves student outcomes. We examine alternative perspectives on the role of assessment and how it is best implemented, review the validity of the most common assessments, and discuss emerging directions in assessment policy and practice. We conclude with implications for policy and research.

Keywords

educational policy, remedial programs, student outcomes, student placement, student placement tests, test validity

Introduction

For most entering community college students, an assessment center is one of the first places they will visit on campus to take examinations that test their proficiency in mathematics, reading, and sometimes writing. According to advice the College Board provides to such students, "You can not 'pass' or 'fail' the placement tests, but it is very important that you do your very best on these tests so that you will have an accurate measure of your academic skills" (College Board, 2011, para. 2). Although it is true that students receive numeric scores on these placement examinations rather than passing or failing grades, 92% of 2-year institutions use the resulting scores for placement

[1]Community College Research Center, Teachers College, Columbia University, New York, NY, USA

Corresponding Author:
Judith Scott-Clayton, Community College Research Center, Teachers College, Columbia University, New York, NY, USA
Email: scott-clayton@tc.columbia.edu

into remedial education (Parsad, Lewis, & Greene, 2003). Often, placement is determined solely on the basis of whether a score is above or below a certain cutoff, although some students may be exempted on the basis of prior ACT, SAT, or high school exit examination scores; students enrolled in noncredit or purely recreational courses may also be exempted. Thus, despite the College Board's reassuring language, placement examination scores are commonly used not merely as a measure of skills, but rather as a high-stakes determinant of students' access to college-level courses.

For the majority of students at community colleges, the consequence of assessment is placement into developmental education.[1] More than half of community college students will eventually enroll in at least one remedial course, and many additional students are assigned to remediation but never enroll (Bailey, 2009; Bailey, Jeong, & Cho, 2010). Estimates of the annual cost of providing remedial instruction "range from about one billion dollars—roughly 1 percent of all public expenditures for postsecondary education . . . —*to three or more times this amount*" (Noble, Schiel, & Sawyer, 2004, p. 300).[2] Additionally, students face the opportunity costs incurred by the extra time that remediation requires, potentially delaying their progress toward a credential.

Yet, despite the prevalence and high costs of remedial assessment and placement, the ultimate benefits of this process are unclear. A number of recent studies on remediation have employed sophisticated designs, such as regression discontinuity and instrumental variable approaches (described later in this review), and found mixed or negative results. Bettinger and Long (2005) found positive effects of mathematics remediation on mathematics credits completed and the probability of transfer for community college students, but they found no significant effects on degree completion and no significant effects of English remediation on any outcome. Bettinger and Long (2009), focusing on students with ACT scores, found positive effects of mathematics and English remediation on bachelor's degree completion, but they also reported some adverse effects of mathematics remediation on the probability of dropping out as well as adverse effects of English remediation on total credits completed. Studies by Calcagno and Long (2008) and Martorell and McFarlin (2011), using broader samples of students, found no impact on most outcomes (including degree completion), with small and mixed effects (both positive and negative) on other outcomes.

Thus, students are assigned to remediation on the basis of assessments, but remediation is not clearly improving outcomes. This calls into question not only the effectiveness of remedial instruction, but also the entire process by which students are assigned to remediation. An analogy can be made to a clinical trial in which individuals' medical histories are assessed to help estimate their ability to benefit from a certain treatment. If the individuals selected for the treatment do not benefit, it could be because the treatment is universally ineffective, because the initial assessment inadequately predicts who is likely to benefit, or because the assessment does not provide enough information to accurately target variations of the treatment to different people. Similarly, if developmental education does not improve outcomes, is it because the "treatment" is broken per se or because the wrong students are being assigned to it? Or is some different or additional treatment required?

This article, drawing on a critical examination of existing literature,[3] broadly examines assessment and placement in community colleges. We first explore whether there is consensus regarding the proper purpose and role of assessment in community colleges. What are the historical, philosophical, and legal contexts surrounding contemporary assessment practices, and how are these policies implemented in practice? Second, we evaluate the research on the most commonly used student assessments. Do the assessments that are currently in use sufficiently predict student outcomes? And, even more importantly, does the use of these assessments seem to improve student outcomes? Finally, we consider whether there are alternative tools that could supplement current assessment and placement procedures, or entirely different models of assessment that might improve outcomes for underprepared students.

The role of assessment—which in this article refers specifically to the assessment of incoming students for determining developmental or college-level placements—deserves attention in a broader discussion of developmental education reform, and we hope that this essay will help illuminate both what is known about the purpose and validity of current assessment strategies and what we still need to learn to design more effective policies. We conclude with a summary of concrete implications for research, policy, and practice.

Purpose and Role of Assessment: Is There Consensus?

Student Assessment and the Community College Open-Door Philosophy

In their comprehensive volume describing community colleges, Cohen and Brawer (2008) discuss "the inevitability of the allocative function" of these institutions (pp. 435-437). In other words, all higher education involves sorting. Students applying to elite and other 4-year institutions are sorted before admission, as colleges accept or reject them according to their test scores and other criteria. Less-advantaged students are sorted after they arrive at open-access institutions. It is the latter group of students that we are concerned with here, along with the testing and placement processes used to sort them into courses with differing levels of difficulty in terms of content and instruction.

There has been significant discussion and debate over whether entry assessments help or harm incoming students, particularly disadvantaged and minority students. As Kingan and Alfred (1993) framed the controversy, assessment can be viewed as a means of tracking and "cooling out" students' college aspirations or as a means of facilitating students' persistence and success; there is support for both views. Students placed in developmental education, particularly at the bottom level, have low odds of eventually moving on to credit coursework. However, the "best practices" literature in developmental education recommends mandatory testing and placement (Boylan, 2002), and the current national trend appears to be toward state standardization of

assessment and enforcement of mandatory placement, suggesting that practitioners and state policymakers believe assessment contributes to students' success.

Historically, the pendulum has swung somewhat in terms of how strictly assessment and placement procedures have been imposed on students. Community colleges from their inception have been open-door institutions and have therefore had to wrestle with the question of how to educate entering students who are unprepared for college-level coursework. From the institutional point of view, the dilemma is framed in terms of the necessity of maintaining academic standards—by controlling entry into college-level courses—in institutions that admit all students (Hadden, 2000). Colleges must maintain standards to establish their legitimacy—to be viewed rightfully as part of the postsecondary sector (Cohen & Brawer, 2008).

For a short period during the 1970s, the mandatory testing, placement, orientation, and course prerequisites fell out of fashion. Proponents of the "student's right to fail" philosophy argued that community college students were adults who should have the freedom to make their own educational decisions, and that this freedom promoted responsibility (Rounds & Andersen, 1985; Zeitlin & Markus, 1996). However, by the end of the decade, these practices were reintroduced as a result of prodding by both legislators and educators concerned with the costs of high failure and dropout rates (Cohen & Brawer, 2008; Rounds & Anderson, 1985).

Challenges were issued almost immediately, and the dilemma became a legal issue. In California, the state's Matriculation Act of 1986 called for improved counseling services and the use of multiple measures in student placement. But the Mexican American Legal Defense and Education Fund (MALDEF) filed a lawsuit on behalf of minority students who claimed they were excluded from courses solely on the basis of placement examinations. The lawsuit was dropped once the community college system chancellor pledged to issue a list of approved tests that were not ethnically or linguistically biased and to fund and enforce the multiple-measures criterion. MALDEF also challenged a state-developed test in Texas (the Texas Academic Skills Program test) as being biased against minority students (Kingan & Alfred, 1993).

Still, a review of this issue in the late 1990s (Fonte, 1997) concluded that the days of a "laissez-fare" approach to developmental education, in which remedial coursework is "voluntary and nondirective" (p. 43), were over. A widely cited compilation of best practices in developmental education stated that mandatory assessment is "a critical initial step in developmental education" that "must be supported by mandatory placement" (Boylan, 2002, pp. 35-36). And a number of studies over the last 15 years have found that community college faculty members and administrators support mandatory assessment and placement (Berger, 1997; Hadden, 2000; Perin, 2006). Faculty members are frustrated when students enroll in courses for which they are not academically prepared; in addition to the resulting challenges for the students, instructors find it challenging to teach a wide range of skill levels within the classroom.

Students would prefer not to be in remediation (Perin, 2006), but if assessment and placement are to be imposed on all students, some observers have emphasized the importance of also providing support services (Bailey et al. 2010; Fonte, 1997; Kingan

& Alfred, 1993; Prince, 2005). College advisors admit that many if not most students take placement tests without understanding their purpose or high-stakes nature (Safran & Visher, 2010). Interviews with community college students have found that they were unprepared for the content and format of the tests, that they were still confused about placement policies after taking the tests, and that many never met with a counselor to discuss their results and subsequent course-taking options (Behringer, 2008; Venezia, Bracco, & Nodine, 2010).

Variation in Assessment and Placement Policies Across States

The brief historical review above demonstrates support among policymakers and educators for an assessment and placement process that places students in courses for which they have the skills to succeed. In the last decade, the debate has evolved to focus on whether institutions can best make these determinations themselves or if the process should be dictated by the state. Arguments for state-standardized assessment and placement policies are that they can establish a common definition of academic proficiency, helping to align secondary and postsecondary academic requirements and expectations; that they can help states measure performance across different colleges and track remedial program effectiveness; and that they can facilitate transfer between colleges (Prince, 2005). Counterarguments cite the importance of institutional autonomy and particularly the importance of institutional freedom to set policies and practices that take into account the particular needs of colleges' local populations (California Community Colleges Chancellor's Office, 2008; Cohen & Brawer, 2008). Institutions may also prefer to select or develop their own assessments and placement procedures due to mistrust of commercially developed products and discomfort with placement determination based on a single test score (Perin, 2006).

Perin's (2006) categories of variation in assessment and placement policy are useful in examining this issue across states. These five categories are as follows: mandatory versus voluntary assessment, type of assessment measure used, whether assessment cutoff scores are set by the state or institution, mandatory versus voluntary placement, and timing of remediation. The last category refers to whether placement into remediation includes a timing requirement, or, as Perin explains, whether developmental education is "a graduation requirement rather than an entry condition" (p. 364). Although Perin's study included only six states (California, Florida, Illinois, New York, Texas, and Washington), she found considerable variation across them. In particular, she found that (a) five of the six states mandated assessment, and in the state that did not, the institutions mandated assessments themselves; (b) a wide variety of assessment instruments were used, and in three states the instrument was determined according to state policy; (c) of those three states, two determined the cut-off scores to be used; (d) remedial placement was required in only four states; and (e) only one state had policy on the timing of remediation, but the individual institutions all had practices that influenced timing. Perin found that some of the state mandates were not strictly implemented in practice.

Several other studies have examined assessment and placement policies across a number of states (Collins, 2008; Jenkins & Boswell, 2002; Prince, 2005; Shults, 2000). The most recent survey of all 50 states was conducted by Ewell, Boeke, and Zis (2008), who asked state-level informants about policies that are intended to improve student transitions through secondary and postsecondary education. One set of questions asked whether the state had a statewide policy on placement, whether a specified set of placement tests is recommended or required, and whether the state sets the cutoff scores for placement. Ewell and his colleagues found that 17 states had a statewide policy governing college placement for all public institutions, while three additional states reported that such a policy was in place for community colleges only. Fourteen states used a common set of placement tests, and an additional state required common tests only in its community colleges. Twelve states determined cutoff scores at the state level, and one additional state mandated specified cutoff scores for community colleges only. The report concludes that the trend is toward more state standardization of assessment and placement.

Indeed, a number of states are actively conducting research to inform consideration of policy change. In 2007, a Task Force on Assessment was established in California to inform statewide discussions on implementing uniform assessment procedures for the state's many (more than 100) community colleges. A survey of the community colleges found that fewer tests were being used than commonly believed; it appears that the institutions were moving in the direction of uniformity themselves (California Community Colleges Chancellor's Office, 2008). Collins (2008) summarized placement policy deliberations and decisions in Virginia, Connecticut, and North Carolina, noting that there were growing internal and external pressures on states to devise "a coherent placement assessment policy framework" (p. 4). Internal pressures include inconsistent entrance standards, alarmingly low student success rates, and unclear course sequences. External pressures come from the national conversations on aligning secondary and postsecondary standards as well as from policymakers' concerns about the costs of such high rates of remediation. For example, a recent joint report from the National Center for Public Policy and Higher Education and the Southern Regional Education Board (2010) recommended "statewide adoption of common assessment practices across broad-access colleges and universities" rather than allowing each school to set its own standards (p. 9).

Centralized policies, while imposing consistency, may have unintended negative consequences. As Cohen and Brawer (2008) have noted, "One of the main problems is the difficulty in setting fixed . . . [standards] for courses and programs that have no set entry requirements" (p. 302). In other words, centrally determined cutoff scores may not appropriately place students within sequences of courses that are institution specific and faculty developed.[4] The movement to standardize placement testing policies does not appear to be linked with a movement to standardize the curricular content of the courses into which students are placed, which would seem to go hand-in-hand with standardizing examinations and cutoff scores. Centralized policies can also negatively affect a state's bottom line. Connecticut's imposition of statewide cutoff scores

resulted in an increase in the number of remedial students, which would increase costs to the students and the state (Collins, 2008).

The implication of this recent policy activity at the state level is that uncertainty underlies current policies and practices—uncertainty about whether the tests and cut-off scores in use is the appropriate ones. Although there remains a great deal of variation within and between states in how assessment is carried out, there is a virtual consensus that it must be done, and the trend is toward increasing state standardization. And although standardization of a fundamentally effective strategy may improve student outcomes, standardization of an ineffective strategy may worsen the situation. Given the assessment strategies in common use, how can we determine whether one test or strategy works better than another? And what evidence is available about the predictive validity of these tests? These questions are addressed in the next section.

Validity of Assessments for Developmental Placement

Commonly Used Placement Exams

The use of placement examinations is nearly universal in community colleges. Parsad et al. (2003) found that 92% of 2-year institutions used placement examination scores for placement into remedial education. Two examinations dominate the market: The ACCUPLACER, developed by the College Board, is used at 62% of community colleges, and the COMPASS, developed by ACT, is used at 46% (Primary Research Group, 2008); note that these percentages are not mutually exclusive, as some schools may "mix and match" depending on the test subject. Although these are the most commonly used tests, several states, including Texas (Evaluation Systems Group of Pearson Education, 2008) and Florida (Common Placement Test Cut-Score Committee, 2006), have worked with testing companies to develop their own examinations.

The ACCUPLACER suite includes a written essay examination as well as computer-adaptive tests in five areas: sentence skills (20 questions), reading comprehension (20 questions), arithmetic (17 questions), elementary algebra (12 questions), and college-level mathematics (20 questions). The tests are not timed, but on average each test takes about 30 min to complete (College Board, 2007, p. 2). Similarly, the COMPASS offers a writing essay as well as untimed computer-adaptive examinations in reading, writing skills, and mathematics. Both ACCUPLACER and COMPASS offer schools the option of including supplementary questions to collect background information on students, such as whether English is the student's first language, whether the student studied algebra in high school, and when the student was last enrolled in a mathematics class.

Manuals published by each vendor (ACT, 2006; College Board, 2003) provide psychometric evidence of test reliability and validity, as well as descriptions of how different score ranges may be interpreted. Yet, both vendors emphasize the importance of performing local validation, preferably every 5 to 7 years, or more frequently if there are changes in course content, examination content, or the characteristics of incoming

students (Morgan & Michaelides, 2005, p. 11). Both vendors offer support services to schools interested in conducting their own analyses. In addition, both vendors suggest that placement decisions may work best when multiple measures are used, not test scores alone (ACT, 2006, p. 2; College Board, 2003, p. A-2).

What Makes an Assessment Valid?

The most recent edition of the *Standards for Educational and Psychological Testing*, produced by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (1999), defines validity as "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests . . . It is the interpretation of test scores required by proposed uses that are evaluated, not the test itself" (p. 9). Similarly, Kane (2006) stated that, "It is not the test that is validated and it is not the test scores that are validated. It is the claims and decisions based on the test results that are validated" (pp. 59-60). This reflects the emphasis in modern validation theory on arguments, decisions, and consequences rather than the mere correspondence of test scores to outcomes (i.e., criteria) of interest. This is what Kane (1992) calls an argument-based approach to validity.

The reference manuals for both major tests follow this approach and identify some of the key assumptions underpinning the validity argument for the use of test scores for course placement. For example, both the COMPASS and ACCUPLACER manuals explain that to be valid, their tests must actually measure what they purport to measure and reliably distinguish between students who are likely or not likely to do well in specific "target" courses; in addition, there should be a positive statistical relationship between test scores and grades in the target courses (ACT, 2006, p. 100; College Board, 2003, p. A-62). Both manuals are explicit, however, that although the above elements are necessary to demonstrate validity, they are not sufficient to demonstrate validity. As the ACCUPLACER manual warns, "ultimately, it is the responsibility of the users of a test to evaluate this evidence to ensure the test is appropriate for the purpose(s) for which it is being used" (College Board, 2003, p. A-62).

What else is required to demonstrate validity? Sawyer and Schiel (2000) explained that for a remedial course placement system to be valid, one must show not only that test scores predict success along the desired dimension, but also that "the remedial course is effective in teaching students the required knowledge and skills" (p. 4). Yet, a persistent fallacy in validity arguments is the idea that test validity can be evaluated without respect to the consequences of how test scores are used, and it would be easy for a consumer of the test manuals to make this mistake. Kane (2006) referred to this fallacy as "begging the question of consequences" (p. 57): Are students with low scores actually likely to benefit from remediation? Simply confirming that a placement examination predicts performance in college-level mathematics does not, on its own, imply that students with low scores should be assigned to remedial mathematics courses. This concept, often overlooked in practice, is central to the actionable

assessment hypothesis—the idea that effective assessments should identify not just who is struggling but also who is likely to benefit from a given treatment. This also makes clear why evaluations of the impact of remediation (or other support services provided on the basis of test scores) are critical to the overall validity of a placement testing system.

Evidence

Do placement tests predict future performance?. The traditional method of measuring predictive validity relies on correlation coefficients, where a coefficient of zero indicates no relationship between the test and the relevant outcome and a coefficient of one indicates perfect predictive power. For example, Armstrong's (2000) study of an unnamed placement examination used at three community colleges in California, as well as Klein and Edelen's (2000) study of the City University of New York's since-abandoned Freshman Skills Assessment Test, rely on correlation coefficients to measure predictive validity. However, correlation coefficients can be insufficiently informative or, even worse, misleading. As the COMPASS manual explains, correlations between mathematics test scores and (for example) grades in college-level mathematics courses are generally computed only for those students who place directly into those courses and not for students whose initial test scores led to their placement in remedial courses. And even if (or indeed, especially if) the test identifies the students who are most likely to succeed, this restriction of the range of variation may decrease the correlation coefficients (ACT, 2006). Moreover, there is no obvious or absolute standard for how large a correlation coefficient should be to be considered sufficiently predictive.

Both ACCUPLACER and COMPASS compute measures of placement accuracy rates, as advocated by Sawyer (1996). Acknowledging that no placement rule can avoid making some mistakes—some students who could have succeeded in the college-level course will be placed into remediation, whereas some students who cannot succeed at the college level will be placed there anyway—this procedure quantifies the percentage of students who are accurately placed into remedial or college-level courses under a given placement rule and definition of success. The first step in computing these rates is to define a measure of success, such as earning a grade of B or higher in college-level mathematics courses. Next, logistic regression is used to estimate the relationship between test scores and the probability of success for those students who score high enough to place into the college-level course. Third, this relationship is extrapolated to students scoring below the cutoff. According to Sawyer (1996), this extrapolation is reasonably accurate as long as no more than 25% of students are assigned to the remedial course. The higher the proportion of students assigned to remediation, the more this procedure must extrapolate a relationship based on a limited subset of students. Finally, the placement accuracy rate is calculated as the sum of "true positives" (i.e., students who are placed at the college level and likely to succeed there) and "true negatives" (i.e., students who are not likely to succeed at the college level and placed into remediation). Students are typically considered likely to succeed

if the estimated probability of success generated by the logistic regression is at least 50% (see, for example, Mattern & Packman, 2009, p. 3).

Evidence regarding placement accuracy rates for the two major testing services can be found in meta-analyses by ACT (2006) for the COMPASS and by Mattern and Packman (2009) of the College Board for the ACCUPLACER. Correlation coefficients (provided only for the ACCUPLACER) are generally higher for the mathematics examinations than for reading and writing assessments. In addition, the coefficients are generally higher when a grade of B or above is used as the success criterion rather than a grade of C or above. Placement accuracy rates (provided for both examinations) generally range between 60% and 80%, and they show less of a pattern across test types and outcome criteria.[5]

Perhaps the most useful measure of validity is provided only in the COMPASS analysis. In addition to detailing placement accuracy rates, the 2006 ACT study also indicates that accuracy rates that result from using the test for placement are typically higher than a baseline accuracy rate that would result if all students were assigned to the standard-level course. Interestingly, results for the COMPASS indicate substantial increases in accuracy rates under the B-or-higher criterion but generally small increases in accuracy rates under the C-or-higher criterion; except for placement into college algebra, using the test with the C-or-higher criterion increased placement accuracy by only 2 to 6 percentage points. Overall, our assessment of the evidence indicates that placement examinations are more useful in mathematics than in reading and writing, and they are more useful for predicting who will perform well in college-level courses than for predicting who will merely pass. It also illustrates how the validity of a test depends on what measure of success one expects it to predict.

Limitations of the existing evidence on predictive validity. Although it is not surprising that the most comprehensive evidence on the predictive power of placement tests comes from the test developers themselves, one might worry about the inherent conflict of interest. As Kane (2006) states,

> It is appropriate (and probably inevitable) that the test developers have a confirmationist bias; they are trying to make the testing program as good as it can be. However, at some point . . . a shift to a more arms-length and critical stance is necessary. (p. 25)

Shifting to this more critical stance, we now examine some limitations of the predictive validity evidence.

First, the validity evidence almost always defines the success criterion as achieving certain minimum grades in the higher level course, but there are limitations to relying on grades as a measure of success. As shown by Bailey et al. (2010), only 30% to 40% of students referred to remediation complete the entire sequence of courses to which they are assigned. Many students never enroll in the course to which they are assigned, and many drop out before a grade is received. Thus, the relationship between test scores and predicted success must be estimated from a restricted sample (those who

would enroll in the course if assigned) and may not be representative of the general population of test takers without stronger assumptions. Beyond this statistical concern, the focus on grades may overlook other important outcomes, such as knowledge acquisition, performance in other courses, persistence, or degree completion. Of course, the COMPASS and ACCUPLACER are not designed to predict these outcomes, and it would be unreasonable to expect a single examination to meet all needs. However, given that these are the predominant tests in use, it is important for policymakers to question whether the success criterion they are meant to predict is the most important one.

Second, placement accuracy rates are themselves estimates, yet the validity studies presented in test manuals provide little basis for evaluating their precision. In other words, how large is the confidence interval around this estimate of placement accuracy? The larger the confidence interval, the more hesitant we should be to rely on the placement accuracy rate when making policy decisions. Sawyer (1996), citing Houston, noted that the precision of the estimate depends upon the proportion of test takers initially assigned to the college-level course. It matters whether 25% or 75% are assigned directly to college-level courses, because it is only for these students that the relationship between test scores and grades in the college-level course can be directly estimated. This relationship then must be extrapolated to students in the rest of the sample, who were initially assigned to remediation. Yet the meta-analysis by ACT (2006) indicated that in only two cases for the COMPASS (the numerical skills test for entrance into arithmetic and the reading skills test for entrance into composition) was the percentage assigned to the college-level course more than 50%. In many cases, the proportion is much lower. Sawyer (1996) suggested that as long as "25% or fewer of the students are assigned to the remedial course, then the procedure described here will estimate the conditional probability of success with reasonable accuracy" (p. 280), but this standard does not appear to be met in most cases. (This information is not available for the ACCUPLACER validity studies).

Third, the evidence on incremental validity is relatively limited. Are these tests better than the alternatives, including assigning all students to the target course, evaluating high school achievement alone, or combining multiple measures for placement decisions? According to a review by Noble et al. (2004), "using multiple measures to determine students' preparedness for college significantly increases placement accuracy . . . For example, test scores and high school grades may be used jointly to identify students who are ready for college-level work" (p. 302). Although the data presented in ACT (2006) show increases in accuracy rates for the COMPASS compared to the predicted rates if all students were assigned to the target course, we could not find similar data for ACCUPLACER. Moreover, comparing the predictive value of the test to using nothing at all (rather than to another method of evaluation) seems a fairly unambitious standard. Even so, the increases in accuracy rates appear to be minimal when a grade of C or higher is used as the success criterion (except for college algebra, for which placement accuracy increases by an estimated 20%).

Finally, as previously mentioned, many schools use assessments in mathematics, reading, and writing not only for placement into developmental courses in those subjects, but also as screens for placement into college-level courses in other subjects more broadly. It is worth noting that the use of COMPASS and ACCUPLACER scores in isolation for placement into college-level science, technology, social science, and other substantive coursework is a type of "off-label" use that has been neither theoretically grounded nor broadly validated.

To summarize, the evidence on the predictive validity of the primary tests currently in use is not as strong as desirable, given the stakes involved; yet this does not necessarily imply that there exists another single test that would be better. Instead, these limitations may represent the limitations of single measures more generally. Improving predictions of future course success may require collecting and effectively using measures beyond a single score on a brief cognitive test—perhaps including noncognitive measures or measures of prior academic experience.

Do better outcomes result when test score cutoffs are used for course placement?. Sawyer (2007) recommended asking: "If we use scores on a particular test to make decisions in the manner recommended . . . will better outcomes result?" (p. 255). In other words, when students are placed into remediation as a result of their test scores, are their later college outcomes (such as grades in the college-level course, persistence, and degree completion) better than they would have been had they gone directly into the college-level course? To answer this question, we need to know about the benefits of correct placement as well as the costs of incorrect placement. This question looks beyond test validity and into the realm of program evaluation. As described above, the effectiveness of remediation is tightly linked to the effectiveness of assessment, yet studies of each have proceeded on parallel tracks, with little to no interaction.

Test validity studies rarely attempt to evaluate whether students' outcomes actually improve as a result of remedial placement, perhaps because doing so is much more complicated than demonstrating a statistical relationship between test scores and outcomes. Because students are not assigned randomly, those assigned to remediation in general would be expected to perform worse than nonremediated students even if remediation were beneficial. Attempting to address this problem, Sawyer and Schiel (2000) used a pretest–posttest approach to evaluate the effectiveness of remediation. Using data from about 2,500 remediated students at 19 colleges, they found that students who were assigned to and completed a remedial course scored significantly higher on the posttest. However, they conceded that a majority of students in their sample never completed the remedial course and acknowledged that without a control group, it is not possible to attribute the entire score gain to the remedial course itself.

To rigorously establish the causal effects of remediation, researchers must identify a source of variation in remedial placement that is unrelated to students' preexisting characteristics, as several economists have recently done with rigorous quasi-experimental research designs. For example, Bettinger and Long (2009) used an instrumental-variables approach with administrative data on 28,000 students pursuing bachelor's degrees in Ohio, taking advantage of the fact that the same test score may lead to different placement

decisions depending on the institution. The authors used the placement rule of the student's nearest college as an instrument for the actual remediation policy they faced.[6] They found that students assigned to remediation were less likely to drop out and more likely to graduate within 6 years.

Less encouraging results come from two other high-quality studies (Calcagno & Long, 2008; Martorell & McFarlin, 2011), both of which used a regression-discontinuity (RD) approach and a broader sample of students (not just those pursuing a bachelor's degree). These RD analyses took advantage of the fact that a student who scores one point below the cutoff is likely to be similar to a student who scores one point above on observed and unobserved dimensions, except that one is assigned to remediation and the other is not. Thus, if students just below the cutoff have significantly higher outcomes than those who score just above, this difference in performance can be attributed to a causal effect of remediation. Calcagno and Long (2008), using Florida data on 68,000 mathematics placements and 24,000 reading placements, found that assignment to remediation increased persistence to the second year and the total number of credits completed but did not increase the completion of college-level credits or the likelihood of completing a degree. Martorell and McFarlin (2011) analyzed data on 445,000 first-time enrollees in Texas and found that assignment to remediation had a negative effect on the number of college-level credits earned as well as negative effects on persistence. They found no effects, positive or negative, on degree completion or eventual labor market outcomes.

Summary. The assessments currently in use at community colleges may be reasonably good at predicting whether students are likely to do well in college-level coursework. Based on meta-analyses conducted by the test makers, both of the major tests currently in use can reasonably be considered valid if the goal is to ensure minimum pass rates in college-level classes. Interestingly, the tests appear to be better at predicting success in mathematics than in English (composition), and they appear to be better at identifying who is likely to earn a B or higher than they are at identifying who is at risk for failure. Incorporating multiple measures may improve this prediction somewhat. Thus, if the ultimate goal of test use is to improve outcomes for low-performing students, the evidence in its favor is far from compelling. Overall, better outcomes do not seem to result for the students who are assigned on the basis of these assessments to remediation, but the costs of remediation are significant for both students and institutions.

The lack of impact could be blamed on the quality of remedial instruction, or perhaps on levels of student preparation that are too low for college-level success with or without remediation. However, Martorell and McFarlin (2011) found little variation in the outcomes of students assigned to remediation across institutions in Texas, which is somewhat surprising given the likely variation in student background, instructor quality, and pedagogy across remedial courses. One possibility is that remedial instruction is uniformly ineffective (or that students are uniformly unable to benefit). An alternative is that the assessments currently in use are focused on predicting only one criterion of success (grades in the college-level course) when other factors may be equally important. The reality may be somewhere in between: Improving assessment may be

a necessary component of improving developmental outcomes, but improvements in assessment alone may be insufficient unless improvements are also made in student preparation and remedial instruction.

Alternative Approaches to Assessment

Our findings above indicate that the common assessments currently in use have some utility but are insufficient in terms of providing enough information to determine the appropriate course of action that will lead to academic progress and success for the vast range of underprepared students. This is likely because students arrive in community colleges underprepared in many ways—not only academically. Conley (2005), among others, has expanded the definition of college readiness beyond academic measures and cognitive strategies to include attitudes and behavioral attributes such as self-monitoring and self-control. Tests such as the COMPASS and ACCUPLACER cannot help community colleges assess whether students might be hampered by the lack of such qualities and, on the basis of those assessments, devise effective interventions.

As noted above, the major test vendors recommend supplementing test scores with other measures for course placement. At least one state, California, requires the use of multiple measures, such as high school transcripts and writing samples, in placing students. The California policy was spurred by the view that a single standardized assessment does a disservice to those from diverse racial and cultural groups; others have made this point and provided evidence for it (Sedlacek, 2004). Given the work of Conley, as well as other support for a more holistic assessment process, does the research literature indicate what additional measures might lead to better placement and student progress, particularly for the community college population?

Alternative or Additional Cognitive Measures

As Safran and Visher (2010) have pointed out, 4-year colleges develop a picture of students' readiness by reviewing transcripts and student work in addition to standardized test scores. Yet community colleges tend to rely on single-test scores for placement in reading, writing, and mathematics. This is likely the reason that we located few studies comparing the outcomes of using one or multiple cognitive measures for incoming community college students.

A small experimental study conducted by Marwick (2004) concluded that the use of multiple measures results in better outcomes than the use of single measures. Marwick randomly assigned students to four alternative mathematics placement procedures: one based on ACCUPLACER scores alone, one based on self-reported high school preparation, one based on the test score and high school mathematics preparation, and one based on student choice. The students assigned to the multiple-measures group—test score and prior mathematics preparation—were less likely to be assigned to remediation but performed no worse in the college-level class than students who

were assigned on the basis of test scores or high school preparation alone. However, the sample included only 304 students from a single community college, and the experimental design and results are not fully described, making it difficult to draw firm conclusions about internal and external validity.

A study of a single California institution found that adding a small number of questions regarding high school history to the computerized assessment increased course placement accuracy, as measured by faculty and student surveys (Gordon, 1999). Another study of students in three large community colleges in California examined whether placement tests or student characteristics predicted performance (in terms of course grades) in three levels of English and mathematics (Armstrong, 2000). The study found that the self-reported high school performance measures were more powerful predictors of student success than the test scores alone, yet the author also found a high degree of variation in grading practices by instructors, pointing out that "misclassification of students," or incorrect placement, may be partly a function of who assigns the grade (p. 691). Another study makes a similar point—that variation in course content within and between community colleges likely makes it difficult to find strong associations between high school grades and test scores and subsequent college performance (Willett, Hayward, & Dahlstrom, 2008). This correlational study, which included data from dozens of California institutions, found modest positive associations between 11th-grade performance in English and mathematics and the level of the first community college course attempted in those disciplines and the grade received.

Although not intended to be used for placement, a new academic diagnostic tool, ACCUPLACER Diagnostics, was recently released by the College Board (2009). The new test is likely a response to criticism that the existing tests, particularly the mathematics test, do not identify the particular content an individual knows or does not know. The new assessment includes English and mathematics tests with five domains per test, and scores are given by test and domain under subheadings of *needs improvement*, *limited proficiency*, and *proficient*. This may represent one step toward a more actionable assessment process.

Noncognitive Measures

Although dictionaries define the word *cognitive* fairly consistently as referring to conscious intellectual activity, the literature reveals many different terms for, or ways to think about, students' noncognitive characteristics. Some have referred to noncognitive characteristics broadly as "students' affective characteristics" (Saxon, Levine-Brown, & Boylan, 2008, p.1). Sedlacek defined noncognitive variables as "variables relating to adjustment, motivation, and student perceptions" (2004, p.7). Conley's (2005) expanded operational definition of college readiness includes four major areas: key cognitive strategies, such as inquisitiveness, analytic skills, and problem-solving abilities; key content knowledge; academic behaviors, such as self-awareness, self-control, study skills, and communications skills; and contextual skills and awareness, including an understanding of the norms and conventions of the postsecondary system.

Although his analysis implies that the first two are cognitive and the latter two are noncognitive, others have categorized critical thinking and reasoning skills as affective skills (Levine-Brown, Bonham, Saxon, & Boylan, 2008).

It is certainly plausible that one's personality and emotional temperament would influence one's academic abilities, and regardless of the variations in language and classification, there is some evidence of an association between affective characteristics and academic performance. Sedlacek (2004) cited numerous studies in support of eight noncognitive variables that may be useful for assessing diverse populations in higher education: positive self-concept, realistic self-appraisal, successful handling of the system (including racism), preference for long-term goals, availability of a strong support person, leadership experience, community involvement, and knowledge acquired in a field. Although a full review of these studies is beyond the scope of this article, Sedlaeck noted that they have found correlations between these noncognitive variables and college grades, retention, and graduation, among other outcomes, particularly for underrepresented minorities.

On the basis of this research, some policymakers and practitioners have called for a more holistic process that would use both cognitive and affective assessments to target remedial coursework as well as other services (e.g., Boylan, 2009), and quite a number of affective assessments exist. Saxon et al. (2008) and Levine-Brown et al. (2008) have provided information on almost three dozen instruments that assess student learning strategies, learning styles, attitudes, study skills, college knowledge, test anxiety, self-efficacy, and personality dimensions, among other variables. Some were developed for particular subpopulations, such as older or minority students. The Learning and Study Strategies Inventory (LASSI), for example, purports to diagnose students' strengths and weakness in 10 different areas including anxiety, attitude, concentration, and time management compared to other students (Saxon et al., 2008).

Yet a 2004-2005 survey of a small sample of 2-year community and technical colleges found that only two of the 29 institutions used noncognitive assessments (Gerlaugh, Thompson, Boylan, & Davis, 2007). Saxon et al. (2008) posited that affective assessments may be infrequently used because institutional decision makers are unaware of the variety and validity of the instruments available. Time and fiscal constraints likely also mitigate against their use, although computerized versions are available.[7] The LASSI website (http://www.hhpublishing.com/_assessments/LASSI/) lists over 2,000 institutions that have administered the test, including secondary and 2-year and 4-year postsecondary institutions, but there is no information available on the scale or regularity of usage within any single institution.

More research on the effectiveness of using multiple measures for academic placement, as well as guidance on the potential uses of the noncognitive assessments, is certainly needed. Do affective assessments, when combined with the scores from the typical assessments, provide information useful for academic placement, particularly for underprepared students? Or are affective assessments more useful in determining which students should be referred to particular campus services, such as mentoring or tutoring? Most colleges offer some innovative models of developmental education,

such as learning communities, accelerated coursework, or the mainstreaming of underprepared students into college courses with extra supports. Because some of these models require additional effort or commitment from students, multiple measures could be useful to colleges in matching students to particular programs.

An interesting related example is the individualized education program (IEP) model that is used to guide the provision of special education supports and services for students with disabilities at the elementary and secondary levels. The IEP model uses a team approach to assess students' academic and personal needs. The IEP team consists of parents, teachers, and other school staff members, who bring together knowledge and experience to design an individualized program that will help the student progress in the general curriculum. Assessment involves tests administered by the team as well as team observations of the student's classroom. It also involves observations from teachers, parents, paraprofessionals, related service providers, administrators, and others. Older students also participate as team members.

Hunter Boylan, director of the National Center for Developmental Education, is among those who have called for this sort of individually targeted approach. Boylan's (2009) model of "targeted intervention for developmental education students" would require generating a comprehensive listing of available campus and community resources, individually assessing each student's skills and characteristics to match them with the specific services likely to be most useful, and providing comprehensive and ongoing advising to help plan, monitor, and revise interventions as needed (p. 15). Although Boylan's model does not necessarily require adding services and may lower some costs by reducing the number of students in remediation, he concedes that it would require a greater investment of both time and money in assessment and individualized advising, which schools may not be able to afford. It is thus unclear whether an IEP-type model is feasible to implement at a broad scale.

Future Directions and Challenges

We now return to our original questions and consider implications for future research and policy. First, there is a fair amount of consensus regarding the role of assessment in community colleges in terms of maintaining open access to the institution while ensuring that students meet minimum standards before proceeding to college-level work. There is much less of a consensus, however, when it comes to determining and implementing assessment and placement policy. From state to state and school to school, there is a high degree of variation in which tests are used, how tests are administered, whether placement recommendations are voluntary or mandatory, and when remediation must be completed. Overall, however, the trend seems to be toward greater standardization of policy at the district or state level.

Second, the student assessments most commonly in use (COMPASS and ACCUPLACER) seem to be reasonably valid predictors of students' grades in college-level coursework, but the placement recommendations that result from the use of these tests do not

clearly improve student outcomes. This suggests a mismatch between the intervention and the assessment that it is based on. Possible responses are to experiment with alternative interventions such as accelerated remediation (e.g., Edgecombe, 2011), or to augment current assessments with additional information that might be used to more closely match students to interventions that will be effective for them.

Third, we find that there are alternative approaches to assessment that have the potential to improve student outcomes. Some evidence suggests that using multiple measures for student assessment and placement—including academic, diagnostic, and affective measures—can provide useful information to institutions that could result in course placement and interventions that better meet students' individual needs. What is likely needed is a new model of actionable assessment that would better identify what students need to be successful while at the same time identifying the level of skills and knowledge that they have at the time of the assessment. The process of implementing a new model of assessment, however, is not without challenges. Colleges may not have the capacity and resources to provide a range of comprehensive assessments or act on the improved information. Particularly in the current economic climate, community colleges may be unable to completely restructure their developmental curriculum; thus, implementing more holistic assessments would be largely fruitless.

The trend toward state standardization of examinations and cutoff scores, as recommended by the National Center for Public Policy and Higher Education and the Southern Regional Education Board (2010), poses another challenge to institutions that may wish to implement more individualized and diagnostic assessment strategies. As discussed above, there are many worthy reasons for such standardization, such as the desire to send more consistent messages to students about college-ready standards and the facilitation of cross-state research on student progress. The current national movement toward common academic standards in the K-12 sector, as exemplified by the Common Core State Standards Initiative (http://www.corestandards.org/), is another effort toward standardization that reflects the same goals. Yet, centrally driven simplifications of the assessment process may work against a more tailored approach, in which colleges might select a range of assessments to guide placement of students into different interventions. And, although the K-12 common core movement includes the establishment of college-ready standards and the allocation of federal funds for the development of new assessment systems, it is unclear how these efforts will be coordinated with the community college assessment frameworks already in place.

Thus, although broad reform of assessment and remedial practices may be necessary, it is unlikely to happen quickly or easily. In the meantime, an increasingly popular trend is simply to assess students earlier. The idea behind early assessment is to offer college placement tests to students in high school, usually in their junior year, as a way of removing the high-stakes context and providing information on skills deficiencies well before college entry. This makes high schools responsible for remediation and may forestall any reform of the tests used by colleges to assess the skills of entering students or the instructional practices used in remedial programs. The

California State University system's Early Assessment Program is just beginning to yield evidence that participation reduces students' need for remediation; a study by Howell, Kurlaender, and Grodsky (2010) found that the program reduced students' probability of needing remediation by roughly four percentage points in mathematics and six percentage points in reading.

Ultimately, our review has uncovered more evidence supporting the need for reform than evidence on what type of reform would work best, but this is not cause for discouragement. Some of the alternatives discussed in the previous section are promising areas for wider implementation and more rigorous evaluation. For example, it would be useful to generate and compare algorithms for placement that combine multiple measures in a way that could be implemented at scale. This might involve comparing the predictive power of placement scores alone to the predictive power of placement scores and affective measures combined, placement scores and high school grades combined, or other combinations. Second, institutions could experiment with using placement test scores (or multiple measures) to place students into alternative treatments. For example, researchers could compare the effectiveness of placement into existing developmental courses and programs versus placement into accelerated courses or placement into regular courses that are augmented by intensive support services or by performance-based scholarships that are contingent on successful student performance. Third, future research should more deeply explore whether current assessment and placement policies have heterogeneous effects. It may be that the current system does work well for some subset of students, but that additional research is needed to identify who those students are. Finally, given the evidence that incoming students are not well informed about assessment and placement policies and practices (e.g., Behringer, 2008; Safran & Visher, 2010; Venezia et al., 2010), there is a need to expand and rigorously evaluate strategies aimed at improving awareness of and preparation for placement examinations.

Acknowledgments

The authors thank Olga Rodriguez and John Wachen, both of the Community College Research Center, for their assistance with the research, and Stanley Rabinowitz of WestEd and Brad Bostian of Central Piedmont Community College (Charlotte, NC) for their useful comments in reviewing an earlier draft.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the Bill & Melinda Gates Foundation.

Notes

1. We use the terms *remedial* and *developmental* interchangeably in this essay
2. Only a small fraction of this amount, perhaps US$6 million per year, is spent on the direct costs of assessments (calculated by multiplying 1.2 million entering students by an average testing cost of about US$5 per student, not including administrative and physical resource costs). The cost per test "unit" ranges from US$1.21 to US$1.66 per student depending on volume, and the typical student takes 3.4 exam units (S. Lewis of ACT, personal communication, May 21, 2010).
3. In addition to reviewing citations from key articles of which we were already aware, we also searched ERIC, Academic Search Premier, Education Full Text, EconLit, JSTOR, ProQuest Digital Dissertations, Google, Google Scholar, and the Teachers College Library for additional references spanning the years from 1990 to 2010. The main search descriptors were: *assessment, ACT, COMPASS, ACCUPLACER, SAT, developmental education, remedial education, placement,* and *tracking*. These descriptors were used in combination with the following terms: *community college, postsecondary, high school, ESL, math, reading, writing, multiple measures, alternative assessment, voluntary, mandatory, effectiveness,* and *validation*. Using these search methods we found thousands of references, which were screened by research assistants. Of these, 106 were found to directly address the research questions. Of these, 60 were initially rated to be highly relevant, as defined by a usefulness rating of at least 2.5 on a scale of 1 to 3. These studies were read closely, and many were ultimately found to be of limited use due to questionable internal validity or narrow external validity (e.g., small nonexperimental studies of school-specific assessments conducted by institutional research staff).
4. The curriculum of the first college-level course is not as standardized as one might imagine, particularly in mathematics. For example, a forthcoming study by Jaggars and Hodara (2011) found that even within a single community college system, options for the first college-level mathematics course range from college algebra, to introductory statistics, to courses examining the history and culture of mathematics.
5. We found only one study by an independent researcher that examined placement accuracy rates. This study, which examined a reading test in place at one community college, found that the test accurately placed students into introductory psychology 77% of the time (Napoli & Wortman, 1995). Note that these authors did not refer to their statistic as a placement accuracy rate but rather as "total cases correctly classified" under a C-or-better success criterion.
6. An instrumental-variables approach can be used when a treatment is not completely randomly assigned but some factor (such as distance to schools with alternative policies) introduces at least some randomness into the process. The approach then seeks to isolate this random variation, separating out the non-random variation due to student ability, preferences, and other factors. In this specific case, the researchers select a sample of students with "marginal" scores that would place them into remediation at some schools but not others. Thus, outcomes for marginal students who live near schools that would place them

into remediation are compared with outcomes for similar students who live near schools that would place them into college-level courses.

7. In a recent telephone interview we conducted as part of a related qualitative study, a Texas community college administrator expressed enthusiasm about the LASSI but said the college can only afford to administer it to a subset of the student population, namely a group of students participating in a program for those who are at risk.

References

ACT. (2006). *COMPASS/ESL reference manual.* Iowa City, IA: Author.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Armstrong, W. B. (2000). The association among student success in courses, placement test scores, student background data, and instructor grading practices. *Community College Journal of Research & Practice, 24*, 681-695. doi:10.1080/10668920050140837

Bailey, T. (2009). Challenge and opportunity: Rethinking the role and function of developmental education in community college. In A. C. Bueschel & A. Venezia (Eds.), *Policies and practices to improve student preparation and success* (New Directions for Community Colleges, No. 145, pp. 11-30). San Francisco, CA: Jossey-Bass. doi:10.1002/cc.352

Bailey, T., Jeong, D. W., & Cho, S. W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review, 29*, 255-270. doi:10.1016/j.econedurev.2009.09.002

Behringer, L. B. (2008). *Remedial education at the community college: A study of student sensemaking.* Unpublished doctoral dissertation, New York University, New York, NY.

Berger, D. M. (1997). Mandatory assessment and placement: The view from an English department. In J. M. Ignash (Ed.), *Implementing effective policies for remedial and developmental education* (New Directions for Community Colleges, No. 100, pp. 33-41) San Francisco, CA: Jossey-Bass. doi:10.1002/cc.10003

Bettinger, E. P., & Long, B. T. (2005). Remediation at the community college: Student participation and outcomes. In C. A. Kozeracki (Ed.), *Responding to the challenges of developmental education* (New Directions for Community Colleges, No. 129, pp. 17-26) San Francisco, CA: Jossey-Bass. doi:10.1002/cc.182

Bettinger, E. P., & Long, B. T. (2009). Addressing the needs of underprepared students in higher education: Does college remediation work? *Journal of Human Resources, 44*, 736-771.

Boylan, H. R. (2002). *What works: Research-based best practices in developmental education.* Boone, NC: Continuous Quality Improvement Network with the National Center for Developmental Education, Appalachian State University.

Boylan, H. R. (2009). Targeted Intervention for Developmental Education Students (T.I.D.E.S). *Journal of Developmental Education, 32*(3), 14-23.

Calcagno, J. C., & Long, B. T. (2008). *The impact of postsecondary remediation using a regression discontinuity approach: Addressing endogenous sorting and noncompliance* (Working Paper 14194). Cambridge, MA: National Bureau of Economic Research.

California Community Colleges Chancellor's Office, Consultation Council Task Force on Assessment. (2008). *Report of the consultation council task force on assessment.* Retrieved from http://www.cccco.edu/Portals/4/Executive/Board/2008_agendas/january/3-5_Assessment%20TF%20Report%2001-08.pdf

Cohen, A. M., & Brawer, F. B. (2008). *The American community college* (5th ed.). San Francisco, CA: Jossey-Bass.

College Board. (2003). *ACCUPLACER OnLine: Technical manual.* New York, NY: Author.

College Board. (2007). *ACCUPLACER coordinator's guide.* New York, NY: Author.

College Board. (2009). *ACCUPLACER Diagnostics.* New York, NY: Author. Retrieved from http://professionals.collegeboard.com/profdownload/accuplacer-diagnostics-brochure.pdf

College Board. (2011). *ACCUPLACER Introduction for students.* Retrieved from http://www.collegeboard.com/student/testing/accuplacer/

Collins, M. L. (2008, July). *It's not about the cut score: Redesigning placement assessment policy to improve student success.* Boston, MA: Jobs for the Future. Retrieved from http://www.jff.org/sites/default/files/cutscoreforWeb_0.pdf

Common Placement Test Cut-Score Committee. (2006, September). *CPT Cut Score Committee final report.* Retrieved from the Florida Department of Education website: http://www.fldoe.org/articulation/pdf/acc_102506ada.pdf

Conley, D. (2005). *College knowledge: What it really takes for students to succeed and what we can do to get them ready.* San Francisco, CA: Jossey-Bass.

Edgecombe, N. (2011, May). *Accelerating the academic achievement of students referred to developmental education* (CCRC Brief No. 55). New York, NY: Community College Research Center, Teachers College, Columbia University. Retrieved from http://ccrc.tc.columbia.edu/Publication.asp?uid=920

Evaluation Systems Group of Pearson Education. (2008). *Texas higher education assessment (THEA) faculty manual* (TA-FM-FACMAN-04 6/08). Amherst, MA: Pearson Education. Retrieved from http://www.thea.nesinc.com/PDFs/THEA_FacultyManual.pdf

Ewell, P., Boeke, M., & Zis, S. (2008). *State policies on student transitions: Results of a fifty-state inventory.* Boulder, CO: National Center for Higher Education Management Systems. Retrieved from http://www.nchems.org/c2sp/documents/TransitionsSurveyReport_April2009.pdf

Fonte, R. (1997). Structured versus laissez-faire open access: Implementation of a proactive strategy. In J. M. Ignash (Ed.), *Implementing effective policies for remedial and developmental education* (New Directions for Community Colleges, No. 100, pp. 43-52) San Francisco, CA: Jossey-Bass. doi:10.1002/cc.10004

Gerlaugh, K., Thompson, L., Boylan, H., & Davis, H. (2007). National Study of Developmental Education II: Baseline data for community colleges. *Research in Developmental Education, 20*(4), 1-4.

Gordon, R. J. (1999, January). *Using computer adaptive testing and multiple measures to ensure that students are placed in courses appropriate for their skills.* Paper presented at the North American Conference on the Learning Paradigm, San Diego, CA (ERIC Document Reproduction Service No. ED425781).

Hadden, C. (2000). The ironies of mandatory placement. *Community College Journal of Research & Practice, 24*, 823-838.

Howell, J. S., Kurlaender, M., & Grodsky, E. (2010). Postsecondary preparation and remediation: Examining the effect of the Early Assessment Program at California State University. *Journal of Policy Analysis and Management, 29*, 726-748.

Jaggars, S. & Hodara, M. (2011, forthcoming). *The opposing forces that shape developmental education: Assessment, placement, and progression at CUNY community colleges* (CCRC Working Paper). New York, NY: Community College Research Center, Teachers College, Columbia University.

Jenkins, D., & Boswell, K. (2002). *State policies on community college remedial education: Findings from a national survey*. Denver, CO: Education Commission of the States (ERIC Document Reproduction Service No. ED470465).

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*, 527-535. doi:10.1037//0033-2909.112.3.527

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, CT: ACE/Praeger.

Kingan, M. E., & Alfred, R. L. (1993). Entry assessment in community colleges: Tracking or facilitating? *Community College Review, 21*(3), 3-16. doi:10.1177/009155219302100302

Klein, S. P., & Edelen, M. O. (2000). *CUNY's testing program, characteristics, results and implications for policy and research*. Santa Monica, CA: RAND.

Levine-Brown, P., Bonham, B. S., Saxon, D. P., & Boylan, H. R. (2008). Affective assessment for developmental students, Part 2. *Research in Developmental Education, 22*(2). Retrieved from http://www.ncde.appstate.edu/resources/reports/documents/RiDE_22_2.pdf

Martorell, P., & McFarlin, I. J. (2011). Help or hindrance? The effects of college remediation on academic and labor market outcomes. *Review of Economics and Statistics, 93*, 436-454. doi:10.1162/REST_a_00098

Marwick, J. D. (2004). Charting a path to success: The association between institutional placement policies and the academic success of Latino students. *Community College Journal of Research & Practice, 28*, 263-280. doi:10.1080/10668920490256444

Mattern, K. D., & Packman, S. (2009). *Predictive validity of ACCUPLACER scores for course placement: A meta-analysis* (Research Report No. 2009-2). New York, NY: College Board. Retrieved from http://professionals.collegeboard.com/profdownload/pdf/09b_765_PredValidity_WEB_091124.pdf

Morgan, D. L., & Michaelides, M. P. (2005). *Setting cut scores for college placement* (Research Report No. 2005-9). New York, NY: College Board. Retrieved from http://professionals.collegeboard.com/profdownload/pdf/060516CBReport2005-9WEB.pdf

Napoli, A. R., and Wortman, P. (1995) Validating college placement tests. *Journal of Applied Research in the Community College, 3*, 143-151.

National Center for Public Policy and Higher Education & Southern Regional Education Board. (2010). *Beyond the rhetoric: Improving college readiness by improving state policy*. Washington, DC: Author.

Noble, J. P., Schiel, J. L., & Sawyer, R. L. (2004). Assessment and college course placement: Matching students with appropriate instruction. In J. E. Wall & G. R. Walz (Eds.), *Measuring up: Assessment issues for teachers, counselors, and administrators* (pp. 297-311). Greensboro, NC: ERIC Counseling and Student Services Clearinghouse and the National Board of Certified Counselors (ERIC Document Reproduction Service No. ED480379).

Parsad, B., Lewis, L., & Greene, B. (2003). *Remedial education at degree-granting postsecondary institutions in fall 2000* (NCES 2004-101). Washington, DC: National Center for Education Statistics.

Perin, D. (2006). Can community colleges protect both access and standards? The problem of remediation. *Teachers College Record, 108*, 339-373.

Primary Research Group. (2008). *Survey of assessment practices in higher education.* New York, NY: Author.

Prince, H. (2005, April). *Standardization vs. flexibility: State policy options on placement testing for developmental education in community colleges* (Achieving the Dream Policy Brief). Boston, MA: Jobs for the Future. Retrieved from http://www.achievingthedream.org/_images/_index03/Policy_brief-StandardFlex.pdf

Rounds, J. C., & Andersen, D. (1985). Placement in remedial college classes: Required vs. recommended. *Community College Review, 13*(1), 20-27. doi:10.1177/009155218501300105

Safran, S., & Visher, M. G. (2010). *Case studies of three community colleges: The policy and practice of assessing and placing students in developmental education courses* (Working Paper). New York, NY: National Center for Postsecondary Research and MDRC. Retrieved from http://www.mdrc.org/publications/548/full.pdf

Sawyer, R. (1996). Decision theory models for validating course placement tests. *Journal of Educational Measurement, 33*, 271-290. doi:10.1111/j.1745-3984.1996.tb00493.x

Sawyer, R. (2007). Indicators of usefulness of test scores. *Applied Measurement in Education, 20*, 255-271. doi:10.1080/08957340701431245

Sawyer, R., & Schiel, J. (2000, April). *Posttesting students to assess the effectiveness of remedial instruction in college.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA (ERIC Document Reproduction Service No. ED442856).

Saxon, D. P., Levine-Brown, P., & Boylan, H. R. (2008). Affective assessment for developmental students, part 1. *Research in Developmental Education, 22*(1). Retrieved from http://www.ncde.appstate.edu/resources/reports/documents/RiDE_22_1.pdf

Sedlacek, W. E. (2004). *Beyond the big test: Noncognitive assessment in higher education* (1st ed.). San Francisco, CA: Jossey-Bass.

Shults, C. (2000). *Institutional policies and practices in remedial education: A national study of community colleges*. Washington, DC: American Association of Community Colleges.

Venezia, A., Bracco, K. R., & Nodine, T. (2010). *One shot deal? Students' perceptions of assessment and course placement in California's community colleges*. San Francisco, CA: WestEd. Retrieved from http://www.wested.org/online_pubs/OneShotDeal.pdf

Willett, T., Hayward, C., & Dahlstrom, E. (2008). *An early alert system for remediation needs of entering community college students: Leveraging the California Standards Test* (Report No. 2007036). Encinitas, CA: California Partnership for Achieving Students Success.

Zeitlin, A. N., & Markus, T. C. (1996). Should remediation be mandatory in the community college? *Community Review, 14*, 27-33.

Bios

Katherine L. Hughes is assistant director for Work and Education Reform Research at the Community College Research Center, Teachers College, Columbia University, New York, NY.

Judith Scott-Clayton is an assistant professor in economics and education, Teachers College, and a senior research associate at the Community College Research Center, Columbia University, New York, NY.