Assessing Developmental Assessment in Community Colleges

Katherine L. Hughes
Judith Scott-Clayton

February 2011

CCRC Working Paper No. 19

A WORKING PAPER IN THE CCRC ASSESSMENT OF EVIDENCE SERIES

Across the first year of a major grant from the Bill & Melinda Gates Foundation, with supplemental funding from Lumina Foundation for Education, CCRC has gathered and synthesized a large body of research evidence regarding strategies that may improve the success of students who attend community college. Working papers in the *Assessment of Evidence Series* use the research literature to draw conclusions and provide evidence-based recommendations in eight major topic areas: developmental assessment, developmental acceleration, developmental mathematics pedagogy, contextualization of basic skills instruction, online learning, non-academic support, institutional and program structure, and organizational improvement. All the papers in the series are made available on CCRC's website (http://ccrc.tc.columbia.edu) as they are released.

Address correspondence to:

Katherine L. Hughes
Assistant Director for Work and Education Reform Research
Community College Research Center
Teachers College, Columbia University
525 West 120th Street, Box 174
New York, NY 10027
212-678-3091
Email: hughes@tc.edu

This research was funded by the Bill & Melinda Gates Foundation. The authors thank Olga Rodriguez and John Wachen, both of CCRC, for their assistance with the research, and Dr. Stanley Rabinowitz of WestEd and Dr. Brad Bostian of Central Piedmont Community College for their useful comments in reviewing an earlier draft.

A previous version of this working paper was released in August 2010.

Abstract

Placement exams are high-stakes assessments that determine many students' college trajectories. More than half of entering students at community colleges are placed into developmental education in at least one subject, based primarily on scores from these assessments, yet recent research fails to find evidence that placement into remediation improves student outcomes. While this has spurred debate about the content and delivery of remedial coursework, another possibility is that the assessment process itself may be broken. In this paper we argue that the debate about remediation policy is incomplete without a fuller understanding of the role of assessment. We then examine 1) the extent of consensus regarding the role of developmental assessment and how it is best implemented, 2) the validity of the most common assessments currently in use, and 3) emerging directions in assessment policy and practice. We conclude with a discussion of gaps in the literature and potential implications for policy and research.

Table of Contents

1. Introduction

For most entering community college students, an assessment center is one of the first places they will visit on campus to take exams testing their proficiency in math, 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."[1] While it is true that students receive numeric scores rather than passing or failing grades, 92% of two-year institutions use the resulting scores for placement into remedial education (Parsad, Lewis, & Greene, 2003).[2] Often, placement is determined solely on the basis of whether a score is above or below a certain cutoff. Thus, despite the College Board's reassuring language, placement exam 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.[3] 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, Jeong, & Cho, 2010; Bailey, 2009). 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 (Phipps, 1998)—to three or more times this amount (Costrell, 1998)" (Noble, Schiel, and Sawyer, 2004, p. 300).[4] Students additionally face the opportunity costs of the extra time that remediation requires, potentially delaying their progress toward a credential.

---

[1] The College Board produces one of the most commonly used placement exams, the ACCUPLACER (http://www.collegeboard.com/student/testing/accuplacer/).

[2] Some students at these schools, however, may be exempted on the basis of prior ACT, SAT, or high school exit exam scores; students enrolled in noncredit or purely recreational courses may also be exempted.

[3] We use the terms "remedial" and "developmental" interchangeably in this essay.

[4] Only a small fraction of this amount, perhaps $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 $5 per student, not including administrative and physical resource costs). Susan Lewis of ACT, Inc., explained in a phone call (May 21, 2010) that the cost per test "unit" ranges from $1.21 to $1.66 per student depending on volume, and that the typical student takes 3.4 exam units.

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 variables approaches (described later in this review), and found mixed or negative results. While Bettinger and Long (2005, 2009) found positive effects of math remediation for younger students, studies by Calcagno and Long (2008) and Martorell and McFarlin (2009) using broader samples of students found no impact on most outcomes (including degree completion), with small mixed positive and negative effects 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 history is assessed in order 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 paper 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 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 deserves attention in a broader discussion of developmental education reform, and we hope that this critical examination of the existing literature[5] will help illuminate both what is known about the purpose and validity of current assessment strategies and what we still need to learn in order to design more effective policies. We conclude with a summary of concrete implications for research, policy, and practice.

2. Purpose and Role of Assessment: Is There Consensus?

2.1 Student Assessment and the Community College Open-Door Philosophy

The purpose of assessment is to sort students into courses whose content and instruction differ in their levels of difficulty.[6] All higher education involves sorting. Students applying to elite and other four-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 students, and the testing and placement process used to sort them, that we are concerned with here.

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) frame the controversy, assessment can be viewed as a means of tracking and "cooling out" students' college aspirations or as a means of facilitating

[5] In addition to citation crawling 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 (for example, small non-experimental studies of school-specific assessments conducted by institutional research staff).

[6] The word "assessment" is often used in the context of learning outcomes or program-level assessments for accreditation. In this paper, we use the word to refer specifically to the assessment of incoming students for determining developmental or college-level placements.

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. On the other hand, 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 so have always 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). But, 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; 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, or TASP, 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," were over. A widely cited compilation of best practices in developmental education states 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 decade and a half have found that community college faculty and administrators support mandatory assessment and placement (Berger, 1997; Hadden, 2000; Perin, 2006). Faculty 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 (Kingan & Alfred, 1993; Fonte, 1997; Prince, 2005; Bailey, Jeong, & Cho, 2010). 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 (Venezia, Bracco, & Nodine, 2010; Behringer, 2008).

2.2 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 facilitate transfer between colleges (Prince, 2005). Counterarguments cite the importance of institutional autonomy and particularly of institutional freedom to set policies and practices that take into account the particular needs of colleges' local populations. In addition, given the discomfort with placement determination based on a single test score, it seems necessary to preserve some institutional flexibility in placement.

Perin's categorization of variation in assessment and placement policy is useful in examining this issue across states. Perin's five categories are: mandatory versus voluntary assessment, type of assessment measure, whether assessment cutoff scores are set by the state or institution, mandatory versus voluntary placement, and timing of remediation (2006). 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). While 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: 1) five of the six states mandated assessment, and in the state that did not, the institutions mandated assessments themselves; 2) a wide variety of assessment instruments were used, and in three states the instrument was determined according to state policy; 3) of those three states, two determined the cut scores to be used; 4) remedial placement was required in only four states; and 5) only one state had policy on the timing of remediation, but the individual institutions all had practices that influenced timing. Some of the state mandates were found to be softened in practice.

Several other studies have examined assessment and placement policies across a number of states (Shults, 2000; Jenkins & Boswell, 2002; Prince, 2005; Collins, 2008). The most recent survey of all 50 states is the 2008 report of the National Center for Higher Education Management Systems Transitions Study (Ewell, Boeke, & Zis, 2008). The study 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 seventeen states have a statewide policy governing college placement for all public institutions, with three additional states reporting that such a policy is in place for community colleges only. Fourteen states use a common set of placement tests, with an additional state requiring common tests only in its community colleges. Twelve states determine cutoff scores at the state level, with one additional state mandating 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 109 community colleges. A survey of the community colleges found that fewer tests were being used than commonly believed; it appears that institutions are moving in the direction of uniformity themselves. Collins (2008) summarizes placement policy deliberations and decisions in Virginia, Connecticut, and North Carolina, noting that there are 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 (NCPPHE) and the Southern Regional Education Board (SREB) recommends "statewide adoption of common assessment practices across broad-access colleges and universities" rather than allowing each school to set its own standards (Shulock, 2010, p. 9).

Centralized policies, while imposing consistency, may have unintended negative consequences. For one, centrally determined cutoff scores may not appropriately place students within sequences of courses that are institution-specific and faculty-developed. 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 exams and cutoff scores. Centralized policies can also negatively impact 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.

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 cutoff scores being used are the appropriate ones. While there remains a great deal of variation—within and between states—in how assessment is done, there is a virtual consensus that it must be done, and the trend is toward increasing state standardization. While 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.

3. Validity of Assessments for Developmental Placement

> Validation involves the evaluation of the proposed interpretations and uses of measurements. … 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. (Kane, 2006, pp. 59–60)

In this section, we describe the two assessments most commonly used at community colleges, discuss what it means for a test to be "valid," and then evaluate these tests' validity using evidence from available research. For those more interested in final conclusions than a detailed discussion, a summary is provided at the section's end.

3.1 Commonly Used Placement Exams

The use of placement exams is nearly universal in community colleges. Parsad, Lewis, and Greene (2003) found that 92% of two-year institutions use placement exam scores for placement into remedial education. Two exams dominate the market: the ACCUPLACER®, developed by the College Board, is used at 62% of community

colleges, and the COMPASS®, developed by ACT, Inc., 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).

The ACCUPLACER suite includes a written essay exam 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 math (20 questions). The College Board also offers ACCUPLACER ESL® and ESL essay exams to assess the English skills of those for whom English is a second language. The tests are not timed, but on average each test takes about 30 minutes to complete (College Board, 2007, p. 2). Similarly, the COMPASS offers a writing essay as well as untimed computer-adaptive exams in reading, writing skills, mathematics, and ESL. Taken together, the COMPASS reading, writing, and math exams typically take about 1.5 to 2 hours to complete. Both ACCUPLACER and COMPASS offer schools the option of including supplementary background questions to collect information 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 math class.

Manuals published by each vendor (College Board, 2003; ACT, Inc., 2006) 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 five to seven years, or more frequently if there are changes in course content, exam content, or incoming students' characteristics (Morgan & Michaelides [College Board], 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.[7]

[7] ACCUPLACER materials emphasize this point, while COMPASS materials merely describe how complementary measures can be collected. For example, the ACCUPLACER manual states: "Also, it should be noted that placement decisions are most accurate when multiple measures are used. When possible, ACCUPLACER scores should be used in conjunction with other available data on student performance" (College Board, 2003, p. A-2). The COMPASS manual states, "To complement the information gathered by the placement-assessment measures described above, the COMPASS system also has available an Educational Planning Form to use in learning more about the student's educational background, needs, plans, and goals" (ACT, Inc., 2006, p. 2).

While these are the most commonly used tests, several states also have worked with testing companies to develop their own exams. For example, Florida uses an adaptation of the ACCUPLACER known as the College Placement Test (CPT Cut Score Committee, 2006), while Texas worked with Pearson Education, Inc., to develop the TASP (Pearson Education, Inc., 2008).

3.2 What Makes an Assessment Valid?

In the most recent edition of the *Standards for Educational and Psychological Testing*, published by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME), validity is defined 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" (1999, p. 9; as cited in Brennan, 2006, p. 2; italics added). This definition and the quotation at the beginning of this section reflect the emphasis in modern validation theory on arguments, decisions, and consequences rather than the mere correspondence of test scores to outcomes (criteria) of interest (see, e.g., Brennan, 2006, pp. 2–3, 8–9). 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. The ACCUPLACER manual explains:

> Although this validation framework acknowledges that validity can never be established absolutely, it requires evidence that (a) the test measures what it claims to measure, (b) the test scores display adequate reliability, and (c) test scores display relationships with other variables in a manner congruent with its predicted properties. (College Board, 2003, p. A-62)

Similarly, the COMPASS manual states:

> Each particular use of test scores needs to be justified by an argument for validity. … The elements of the validity argument supporting this use include the following:

> - The COMPASS tests measure the skills and knowledge students need to succeed in specific courses.
> - Students who have the skills and knowledge necessary to succeed in specific courses are likely to perform satisfactorily on the COMPASS tests, and students without those skills are not.
> - Higher levels of proficiency on the COMPASS tests are related to higher levels of satisfactory performance in the course.
>
> If course placement is a valid use of these tests, then a significant, positive statistical relationship between COMPASS test scores and course grades would be expected. (ACT, Inc., 2006, p. 100)

Both passages suggest that the above elements are *necessary* to demonstrate validity but are careful not to claim they are *sufficient* to demonstrate validity.[8] The ACCUPLACER manual directly states the limited nature of its own validity evidence:

> In addition, it should be noted that although test developers must provide evidence to support the validity of the interpretations that are likely to be made from test scores, 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) of ACT, Inc., explain that for a remedial course placement system to be valid, one must show not only that test scores are predictive of 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) refers to this fallacy as "begging the question of consequences" and provides this example (based on Cronbach & Snow, 1977):

> Assume, for example, that a test is an excellent predictor of performance in two treatment options, A and B, but that … everyone does uniformly better in treatment A than in

[8] Richard Sawyer of ACT, Inc., has written that "accurately classifying students … is necessary, but not sufficient, for a placement system as a whole to be effective" (1996, p. 272).

> treatment B. In this case the test scores are not in themselves at all useful for placement decisions; the optimal policy is to assign everyone to treatment A. The validity of the interpretation [of the test score] as a predictor of performance in the two treatment options does not support the validity of the proposed placement decisions. (p. 57)

This example could very well describe the context of developmental assessment (except that if "college-level coursework" is considered as treatment A, it is not that students do uniformly better when assigned directly to college level but rather that they often do no worse). Simply confirming that a placement exam predicts performance in college-level math does not, on its own, imply that students with low scores should be assigned to remedial math. Although it may be beyond the domain of test developers, an important component of the validity argument is whether students with particular scores are likely to perform better in one course than another. This component, 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.

3.3 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, both Armstrong's (2000) study of an unnamed placement exam in use at three community colleges in California and Klein and Edelen's (2000) study of CUNY's since-abandoned Freshman Skills Assessment Test rely on correlation coefficients to measure predictive validity.

But correlation coefficients can be insufficiently informative or, even worse, misleading. As the COMPASS manual explains, correlations between math test scores and (for example) grades in college-level math are generally computed only for those students who place into college-level math, and even if (or indeed, especially if) the test

identifies the students most likely to succeed, this restriction of the range of variation may decrease the correlation coefficients (ACT, Inc., 2006, p. 101). 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).[9] 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, while some students who cannot succeed at the college level will be placed there anyway—this procedure quantifies what percentage of students are accurately placed into remediation 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 math. 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.[10] Finally, for different placement rules (which may involve only a test score or may involve multiple measures), the placement accuracy rate is calculated as the sum of "true positives"—students who are placed at the college level and likely to succeed there—and "true negatives"—students who are not likely to succeed at the college level and placed into remediation.[11]

A summary of the evidence on placement accuracy rates for the two major testing services is provided in Table 1, based on a meta-analysis by ACT, Inc., (2006) for the COMPASS and a meta-analysis by Mattern and Packman (2009) of the College Board for the ACCUPLACER. Only results for analyses based on at least 10 schools are shown in the table. The COMPASS studies are divided by specific target courses (that is, the courses students would be assigned to based on a passing score), while the ACCUPLACER studies are aggregated across multiple target courses linked to each

[9] Note, however, that the ACCUPLACER studies place far more emphasis on traditional measures of correlation coefficients.

[10] According to Sawyer (1996), this extrapolation is reasonably accurate as long as no more than 25 percent 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.

[11] Students are typically considered likely to succeed if the estimated probability of success generated by the logistic regression is at least 50 percent (see, e.g., Mattern & Packman, 2009, p. 3).

Table 1
Placement Accuracy Rates for COMPASS and ACCUPLACER

Test	Target Course	Success Criterion: B or Higher				Success Criterion: C or Higher			
		Number of Colleges	Correlation Coefficient	Average/ Median Accuracy Rate	Median Increase in Accuracy Rate	Number of Colleges	Correlation Coefficient	Average/ Median Accuracy Rate	Median Increase in Accuracy Rate
COMPASS[a]									
Writing Skills	Composition	68	—	66	19	39	—	67	2
Reading Skills	Composition	28	—	60	10	12	—	67	2
Reading Skills	Psychology	11	—	68	31	9	—	67	4
Numerical Skills/Pre-algebra	Arithmetic	26	—	70	16	16	—	72	4
Numerical Skills/Pre-algebra	Elementary Algebra	38	—	67	25	24	—	63	6
Algebra	Intermediate Algebra	29	—	71	25	17	—	68	5
Algebra	College Algebra	23	—	72	43	19	—	67	20
ACCUPLACER[b]									
Sentence Skills	Composition, Reading	21	0.19	59	—	21	0.13	75	—
Reading Comprehension	Composition, Reading	25	0.17	62	—	25	0.10	80	—
Arithmetic	Basic Math to Precalculus	13	0.29	66	—	13	0.23	84	—
Elementary Algebra	Basic Math to Precalculus	34	0.27	65	—	34	0.25	73	—

Note. The analyses above are based on more than 10 schools. Increases in accuracy rates are calculated by comparing the predicted accuracy rates under the given placement rule to the predicted accuracy rate if all students were placed in the target standard-level course. A dash indicates that data are unavailable.

[a]Source: ACT, Inc., 2006, pp. 103–104. [b]Source: Mattern and Packman, 2009, p. 4.

exam. Both meta-analyses evaluate accuracy rates under two definitions of success: earning a B or higher in the target course and earning a C or higher. With a B-or-higher criterion, placement accuracy rates range from 60–72% for the COMPASS exams and 59–66% for the ACCUPLACER exams. With a C-or-higher criterion, placement accuracy rates range from 63–72% for the COMPASS and 73–84% for the ACCUPLACER.

The ACT, Inc., (2006) analysis also indicates the typical increase in accuracy rates that results from using the test for placement (compared with assigning all students to the standard-level course). This is a means of evaluating incremental validity, or how much prediction is improved by using the test. Interestingly, results 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 for the COMPASS (except for placement into college algebra, using the test with the C-or-higher criterion increased placement accuracy by only 2–6 percentage points). This implies that COMPASS exams 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. This information was not provided in the ACCUPLACER study.

Limitations of the existing evidence on predictive validity. While 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, especially for high-stakes testing programs, a shift to a more arms-length and critical stance is necessary in order to provide a convincing evaluation of the proposed interpretations and uses. (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 in Bailey, Jeong, and Cho (2010), only 30–40% 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 exam to meet all needs. But, 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. (ACT, Inc., also offers other types of assessments, which are discussed in section 4.)

Second, placement accuracy rates are themselves estimates, yet the validity studies presented in test manuals provide little basis for evaluating their precision. Sawyer (1996), citing Houston (1993), notes that precision depends upon the proportion of test-takers assigned to the lower-level course. Since the relationship between test scores and grades in the higher-level course must be estimated using data from only those who score above the cutoff and then extrapolated to those below, it matters whether 25% score below the cutoff or 75% do. Yet only in 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 lower-level course less than 50%. In many cases, the proportion is much higher. Sawyer (1996) suggests 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, which focus more attention on traditional correlation coefficients).

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, Schiel, and Sawyer (2004), "Using multiple measures to determine students' preparedness for college significantly increases placement accuracy (ACT, 1997; Gordon, 1999; Roueche & Roueche, 1999). For example, test scores and high school grades may be used jointly to identify students who are ready for college-level work" (p. 302). While Table 1 shows 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 the use of the algebra test increases placement accuracy by an estimated 20 percentage points).

Finally, as previously mentioned, many schools use math and reading/writing assessments 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 additional noncognitive measures or broader measures of prior academic experience and outcomes.

Do better outcomes result when test score cutoffs are used for course placement? Sawyer (2007) recommends asking: "If we use scores on a particular test to make decisions in the manner recommended … will better outcomes result?" (p. 255). 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 benefit overall from the remedial placements that result—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 non-remediated students even if remediation were beneficial (if the test is valid for that purpose). Controlling for preexisting demographic and academic characteristics improves upon naïve comparisons of these two groups but does not eliminate the possibility of preexisting differences on unobserved dimensions.

In order to 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) use 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 upon the institution. The authors use the placement rule of the student's nearest college as an instrument for the actual remediation policy they faced.[12] They found that students assigned to remediation are less likely to drop out and more likely to graduate within six years.

Less encouraging results come from two other high-quality studies (Martorell & McFarlin, 2009; Calcagno & Long, 2008), both of which use a regression-discontinuity (RD) approach and a broader sample of students (not just those pursuing a BA). These

[12] An instrumental-variables approach can be used when a treatment is not completely randomly assigned but some factor or "instrument" (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, etc. 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.

RD analyses take 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 state data on 68,000 math placements and 24,000 reading placements, found that assignment to remediation increases persistence to the second year and the total number of credits completed but does not increase the completion of *college-level* credits or the likelihood of completing a degree. Martorell and McFarlin (2009) analyzed data on 445,000 first-time enrollees in Texas and found that assignment to remediation has 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.

Finally, Sawyer and Schiel (2000) used a pre-test–post-test approach to evaluate the effectiveness of remediation. Using data from about 2,500 remediated students at 19 colleges, they found that students who are assigned to and complete a remedial course score significantly higher on the post-test, suggesting significant knowledge gains. However, they concede that the majority of students in their sample never completed the remedial course and acknowledge that not all of the test score gains may be attributable to the remedial course itself.

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 the evidence presented in Table 1, 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 math 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 (2009) 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 to identify. The reality may be somewhere in between: improving assessment may be a necessary component of improving developmental outcomes but may not be sufficient unless corresponding improvements are made in student preparation and remedial instruction.

4. 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. David 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 so that they may 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 disserves those from diverse racial and cultural groups; others have made this point and provided evidence for it (Sedlacek, 2004). Given the work of Conley and 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?

4.1 Alternative or Additional Cognitive Measures

As Safran and Visher (2010) point out, four-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 math. 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 math 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 math preparation, and one based on student choice. The students assigned to the "multiple measures" group—test score and prior math—were less likely to be assigned to remediation but performed no worse in the college-level class than students who were assigned based on 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 the study's 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 course grades in three levels of

English and math (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. 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 particular correlational study, which included data from dozens of California institutions, found modest positive associations between 11th-grade performance in English and math and the level of the first community college course attempted in those disciplines and grade received.

While not intended to be used for placement, a new academic diagnostic tool, ACCUPLACER Diagnostics, was recently released by the College Board. The new test is likely a response to criticism that the existing tests—particularly the math test—do not identify the particular content an individual knows or does not know. The new assessment includes English and math tests with five domains per test, and scores are given by test and domain, under subheadings of "needs improvement," "limited proficiency," and "proficient." The College Board recommends using Diagnostics in high school as a pre- and post-test tool to assess academic progress, to prepare for placement tests, or even after placement tests to better identify areas of strengths and weaknesses. This is perhaps one step toward a more actionable assessment process.

4.2 Noncognitive Measures

While dictionaries define "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 refer to noncognitive characteristics broadly as "students' affective characteristics" (Saxon, Levine-Brown, & Boylan, 2008, p. 1). Sedlacek defines 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. While his analysis implies that the first two are cognitive and the latter two are noncognitive, others categorize 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) cites 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, successfully handling the system (racism), preference for long-term goals, availability of a strong support person, leadership experience, community involvement, and knowledge acquired in a field. While a full review of these studies is beyond the scope of this paper, they have found correlations between these noncognitive variables and college grades, retention, and graduation, among other outcomes, particularly for underrepresented minorities. Schunk (1984) reviewed many studies of self-efficacy (one's own judgment of one's capabilities) in elementary school children and found that it influences academic persistence and performance.

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 (see, e.g., Boylan, 2009). Yet, a 2004–05 survey of a small sample of two-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) posit 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 impede the use of affective assessments, although computerized versions are available. Saxon et al. (2008) and Levine-Brown et al. (2008) provide 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 of students, such as adults 25 and older or minority students.

There is certainly a need for more research on the effectiveness of using multiple measures for academic placement, as well as guidance on the potential uses of the noncognitive assessments. Do affective assessments provide information useful for academic placement, when combined with the scores from the typical assessments, 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. Since 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, who bring together knowledge and experience to design an individualized program that will help the student progress in the general curriculum. Assessment involves examination by the team of the student's classroom and other tests, as well as 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" (T.I.D.E.S.) would require (p. 15):

- taking an inventory of available campus and community courses and services,

- developing student profiles to determine the types of services that might be helpful to students with various characteristics,
- assessing individual students' skills and characteristics,
- advising students using this assessment information to plan interventions,
- delivering targeted interventions according to the plan,
- monitoring students and evaluating their progress, and
- revising the targeted interventions as necessary.

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 (p. 20). It is thus unclear whether an IEP-type model is feasible to implement for all incoming community college students, or even some subset.

5. 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 upon. Possible responses are to experiment with alternative interventions (such as accelerated remediation, the topic of another paper in CCRC's series on developmental education reform) 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 in addition to 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 likely lack the ability to conduct wholesale restructuring of their developmental curricular offerings, so implementing more holistic assessments would be largely fruitless.

The trend toward state standardization of examinations and cutoff scores, as recommended by NCPPHE and SREB (2010), poses another challenge to institutions that may wish to implement more individualized and diagnostic assessment strategies. As discussed, 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 (i.e., the Common Core State Standards Initiative) 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, while the K-12 common core

movement includes the setting 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, while 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 give students the assessments earlier. The idea behind early assessment strategies is to offer college placement tests to students in high school, usually in their junior year, to remove the high-stakes context and provide information on skills deficiencies well before college entry. This makes high schools responsible for remediation and may forestall any reform of the college tests or instruction. The California State University system's Early Assessment Program is just beginning to yield evidence that participation reduces students' probability of needing remediation; a study by Howell, Kurlaender, and Grodsky (2010) found the program reduced students' probability of needing remediation by roughly four percentage points in math 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 test algorithms for placement that combine multiple measures of preparedness in a way that could be implemented consistently and at scale. This might involve comparing the usefulness of placement scores alone to combinations of academic scores plus a selection of affective measures, test scores plus high school grades in academic subjects, and other combinations of traditional and alternative measures. Second, institutions could experiment with using placement tests (or multiple measures) for targeting of alternative treatments, enabling researchers to compare the effectiveness of placement into existing developmental levels versus placement into accelerated courses or placement into regular courses plus intensive support services or performance-based payments. 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 we need to do a better job of

identifying who those students are. Finally, given the evidence that incoming students are not well informed about assessment and placement policies and practices, there is a need to expand and rigorously evaluate strategies aimed at improving awareness of and preparation for placement exams.

While the field has yet to reach a consensus regarding the best directions for assessment reform, we do see consensus around the need for change in order to drastically improve persistence and graduation rates. Of course, improving assessment is only one facet of a broader agenda for reforming developmental education, but since students' first experiences with community colleges are with the assessment and placement process, this is as good a place as any to begin.

References

ACT, Inc. (1997). *ACT assessment technical manual*. Iowa City, IA: Author.

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

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*(8), 681–695.

Bailey, T. (2009). Challenge and opportunity: Rethinking the role and function of developmental education in community college. *New Directions for Community Colleges, 145*, 11–30.

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*(2), 255–270.

Behringer, L. B. (2008). *Remedial education at the community college: A study of student sensemaking* (Doctoral dissertation). New York, NY: New York University Steinhardt School of Culture, Education, and Human Development.

Berger, D. M. (1997). Mandatory assessment and placement: The view from an English department. *New Directions for Community Colleges, 100*, 33–41.

Bettinger, E. P., & Long, B. T. (2005). Remediation at the community college: Student participation and outcomes. *New Directions for Community Colleges, 129*, 17–26.

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*(3), 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.

Brennan, R. L. (Ed.). (2006). *Educational measurement* (4th ed.). Westport, CT: ACE/Praeger Publishers.

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

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: College Board.

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

Collins, M. L. (2008). *It's not about the cut score: Redesigning placement assessment policy to improve student success*. Boston, MA: Jobs for the Future.

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.

CPT Cut Score Committee. (2006). *CPT cut score committee final report.* Retrieved from Florida Department of Education website: http://www.fldoe.org/articulation/pdf/acc_102506ada.pdf

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York, NY: Irvington Publishers.

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.

Fonte, R. (1997). Structured versus laissez-faire open access: Implementation of a proactive strategy. *New Directions for Community Colleges, 100,* 43–52.

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.

Hadden, C. (2000). The ironies of mandatory placement. *Community College Journal of Research & Practice, 24*(10), 823–838.

Houston, W. M. (1993, April). *Accuracy of validity indices for course placement systems.* Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA.

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*(4), 726–748.

Jenkins, D., & Boswell, K. (2002). *State policies on community college remedial education: Findings from a national survey*. Denver, CO: Education Commission of the States.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*(3), 527–535.

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

Keigher, A. (2009). *Characteristics of public, private, and Bureau of Indian Education elementary and secondary schools in the United States: Results from the 2007-08 schools and staffing survey* (No. NCES 2009-321). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

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

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).

Martorell, P., & McFarlin, I. J. (2009). *Help or hindrance? The effects of college remediation on academic and labor market outcomes.* Unpublished manuscript, RAND and University of Michigan.

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*(3), 263–280.

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.

Morgan, D. L., & Michaelides, M. P. (2005). *Setting cut scores for college placement* (Research Report No. 2005-9). New York, NY: College Board.

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 & Student Services Clearinghouse and the National Board of Certified Counselors.

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

Pearson Education, Inc. (2008). *Texas Higher Education Assessment (THEA) faculty manual: A guide to THEA test results* (No. TA-FM-FACMAN-04). Amherst, MA: Author.

Perin, D. (2006). Can community colleges protect both access and standards? The problem of remediation. *Teachers College Record, 108*(3), 339–373.

Phipps, R. (1998). *College remediation: What it is, what it costs, what's at stake*. Washington, DC: Institute for Higher Education Policy.

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

Prince, H. (2005). *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.

Roueche, J. E., & Roueche, S. D. (1999). *High stakes, high performance: Making remedial education work.* Washington, DC: Community College Press.

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

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.

Sawyer, R. (1996). Decision theory models for validating course placement tests. *Journal of Educational Measurement, 33*(3), 271–290.

Sawyer, R. (2007). Indicators of usefulness of test scores. *Applied Measurement in Education, 20*(3), 255–271.

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.

Saxon, D. P., Levine-Brown, P., & Boylan, H. R. (2008). Affective assessment for developmental students, part 1. *Research in Developmental Education, 22*(1), 1–4.

Schunk, D. H. (1984). Self-efficacy perspective on achievement behavior. *Educational Psychologist, 19*(1), 48–58.

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

Shulock, N. (2010). *Beyond the rhetoric: Improving college readiness by improving state policy*. Washington, DC: National Center for Public Policy and Higher Education & Southern Regional Education Board. Retrieved from http://publications.sreb.org/2010/Beyond%20the%20Rhetoric.pdf

Shults, C. (2000). *Institutional policies and practices in remedial education: A national study of community colleges*. Washington, DC: American Association of Community Colleges. Retrieved from http://www.eric.ed.gov/PDFS/ED447884.pdf

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.

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 (Cal-PASS).

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