

CCRC BRIEF

NUMBER 50

FEBRUARY 2011

Assessing Developmental Assessment in Community Colleges

Katherine L. Hughes
Judith Scott-Clayton

Most community college students begin their college careers at an assessment center, taking exams to test their proficiency in math, reading, and sometimes writing—and for most, the consequence of assessment is placement into developmental education. More than half of community college students enroll in at least one remedial course, and many additional students are assigned to remediation but never enroll (Bailey, 2009; Bailey, Jeong, & Cho, 2010).

Despite its ubiquity, the benefits of remediation (or “developmental education”; we use these terms interchangeably) are unclear. A number of recent studies with sophisticated designs have found little evidence that remediation has positive effects on student outcomes. This could mean that remedial instruction is ineffective—or it could mean that assessments fail to identify which students could benefit from remediation. This calls into question not only the effectiveness of remedial instruction but also the entire process by which students are assigned to remediation. This Brief, based on a longer paper, reviews the literature on assessment and placement to examine the role of assessment in community colleges, evidence on the effectiveness of commonly used placement exams, and potential alternatives to traditional modes of assessment.

The Purpose and Role of Assessment

The purpose of assessment is to sort students into courses of varying difficulty. While students applying to selective institutions are sorted before admission, students who enroll in community colleges are sorted afterward. As open-door institutions, community colleges endeavor to educate underprepared students while simultaneously maintaining their academic standards (thus preserving their status as legitimate members of the postsecondary sector). Assessment and placement practices can help community colleges reconcile these two sometimes conflicting aims.

While there was once greater debate over how strictly to impose assessment and placement procedures on students, mandatory testing and placement is now commonly recommended in the literature on best practices in developmental education (Boylan, 2002) and by many community college faculty and administrators (Berger, 1997; Hadden, 2000; Perin, 2006). There is, however, less of a consensus when it comes to determining and implementing

the details of assessment and placement policy, and the debate over student placement has evolved to focus on whether institutions can best make placement determinations themselves or if the process should be dictated by the state. A study by Perin (2006) examined the policies of six different states along a variety of dimensions: 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—i.e., 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 that their policies varied considerably.

Proponents of state-standardized assessment and placement policies argue that it can (1) establish a common definition of academic proficiency, helping to align secondary and postsecondary academic requirements; (2) help states measure performance across different colleges and track remedial program effectiveness; and (3) facilitate transfer between colleges (Prince, 2005). Counterarguments cite the importance of institutional autonomy and freedom to set policies that take into account the particular needs of colleges’ local populations. And given the discomfort with placement determination based on a single test score, it seems necessary to preserve some institutional flexibility in placement. Still, it appears that the current national trend is toward state standardization of assessment and enforcement of mandatory placement (Ewell, Boeke, & Zis, 2008).

While standardization of an effective strategy may improve student outcomes, standardization of an ineffective strategy may worsen them. Research on assessment and placement can yield vital information about whether one test or strategy works better than another and how well tests can predict student performance in related courses.

Validity of Assessments for Developmental Placement

The use of placement exams is nearly universal in community colleges. Parsad, Lewis, and Greene (2003) found that 92% of two-year institutions use exam scores for placement into remedial education. While a few states have worked with testing companies to develop their own exams, 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, is used at 46% (Primary Research Group, 2008; 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 and computer-adaptive tests in five areas: sentence skills,

reading comprehension, arithmetic, elementary algebra, and college-level math. The College Board also offers ACCUPLACER ESL and ESL essay exams. The tests are not timed, but on average each test takes about 30 minutes to complete (College Board, 2007). Similarly, the COMPASS offers a writing essay and untimed computer-adaptive exams in reading, writing skills, mathematics, and ESL. Taken together, the COMPASS reading, writing, and math exams typically take 1.5–2 hours to complete. Both ACCUPLACER and COMPASS offer schools the option of including supplementary background questions for 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 math class.

Both the College Board (2003) and ACT (2006) provide evidence of test reliability and validity and descriptions of how different score ranges may be interpreted. Yet both vendors emphasize the importance of performing local validation every five to seven years, or more frequently if there are changes in course content, exam content, or incoming students' characteristics—and both offer support services to schools interested in conducting their own analyses (Morgan & Michaelides, 2005). In addition, both suggest that placement decisions may work best when multiple measures are used. A correlation between scores on a placement exam and performance in a specific course may be necessary to consider a test valid, but it is not sufficient.

For a course placement system to be valid, there must be evidence that test scores are predictive of success along the desired dimension and that “the remedial course is effective in teaching students the required knowledge and skills” (Sawyer & Schiel, 2000, p. 4). 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. Test validity cannot be evaluated without respect to the consequences of how test scores are used.

Meta-analyses of placement outcomes for the COMPASS (ACT, 2006) and ACCUPLACER (Mattern & Packman, 2009) evaluate correlation coefficients and placement accuracy (the likelihood that a given student is correctly placed in either the remedial or college-level course) under two definitions of success: earning a B or higher in the target course and earning a C or higher. In general, correlation coefficients (available only for the ACCUPLACER) are higher for math exams than for reading/writing and higher for the B-or-higher criterion than for the C-or-higher criterion. Placement accuracy rates for both exams generally range from 60–80% and show less of a pattern across test types and outcome criteria.

The ACT (2006) analysis provides an additional interesting measure of validity—the typical increase in accuracy rates that results from using the test for placement instead of assigning all students to the standard-level course. 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. Overall, the evidence indicates that placement exams are more useful in math than in reading/writing and 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.

Alternative Approaches to Assessment

While the COMPASS and ACCUPLACER exams have some predictive value, incorporating multiple measures may improve these predictions further. Traditional assessments may not provide sufficient information to determine which course of action will lead to academic success 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. Additional measures of student preparedness would lead to a more holistic assessment process—and potentially to better placement recommendations and outcomes for community college students.

Cognitive Measures

Since most community colleges rely on single test scores for placement in reading, writing, and math, there are few studies comparing the outcomes of using multiple cognitive measures for incoming community college students. Still, some evidence suggests that using additional cognitive measures can improve placement accuracy.

Marwick (2004) randomly assigned students to math placement procedure conditions based on: (1) ACCUPLACER scores alone, (2) self-reported high school preparation, (3) ACCUPLACER scores and high school math preparation, and (4) student choice. Students in the “multiple measures” group—test scores and prior math experience—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, since the sample included only 304 students from a single community college, and the experimental design and results are not fully described, it is difficult to draw firm conclusions about the study's internal and external validity. Another study found that adding questions about students' high school experience to the computerized assessment increased course placement accuracy (Gordon, 1999). In addition, Armstrong's (2000) study found that self-reported high school performance measures predicted student success in math and English better than test scores alone. However, Armstrong also found a high degree of variation in grading practices, pointing out that incorrect placement may be partly a function of who assigns the grade.

Perhaps in response to criticism that existing tests do not provide enough information about students' content knowledge, the College Board released a new diagnostic tool, ACCUPLACER Diagnostics. Recommended for use in high schools, Diagnostics provides an analysis of students' strengths and weaknesses in English and math and may be useful in preparing students for college placement tests.

Noncognitive Measures

There is some evidence of an association between noncognitive (or “affective”) characteristics and academic performance. Sedlacek (2004) identifies eight noncognitive variables that may be useful for assessing diverse

populations: 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. Studies have found correlations between these variables and college grades, retention, and graduation, 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.

While some policymakers and practitioners have called for an assessment process that would incorporate both cognitive and affective measures (Boylan, 2009), a 2004–05 study of two-year community and technical colleges found that only two of the 29 institutions surveyed used noncognitive assessments (Gerlaugh, Thompson, Boylan, & Davis, 2007). Saxon, Levine-Brown, and Boylan (2008) posit that institutional decision makers may be unaware of the variety and validity of the instruments available. Time and fiscal constraints likely also impede the use of affective assessments.

There is 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. Are affective assessments useful for academic placement, or are they more useful for referring students to campus services? Most colleges offer innovative models of developmental education, such as learning communities or accelerated coursework, and multiple measures could be useful in matching students to particular programs.

This sort of individually targeted approach is evident in 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. With the IEP model, a team of parents, teachers, and other school staff assess students' academic and personal needs. Hunter Boylan has also advocated an individually targeted approach. Boylan's (2009) model of "targeted intervention for developmental education students" (T.I.D.E.S.) involves assessing students' skills and characteristics, using information from assessments to provide high-quality advising, and matching students to interventions that are likely to improve their outcomes. While this model could eventually lower some costs by reducing the number of students in remediation, many schools may lack the time and money necessary for detailed assessment and individualized advising.

Conclusion and Recommendations

This review has found evidence of a range of assessment and placement policies in place. 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.

The student assessments most commonly in use (COMPASS and ACCUPLACER) seem to be reasonably valid predictors of students' grades in college-level courses, at least in math when students are expected to earn a B or higher in the target course. Validity appears weaker for the reading/writing exams and for predicting who will earn a C or higher. And importantly, the placement recommendations

generated by these tests do not clearly improve student outcomes. This suggests a mismatch between interventions and assessments.

Alternative approaches to assessment may have the potential to improve student outcomes. 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. Colleges likely need a new model of "actionable assessment" that would better identify what students need to be successful in addition to identifying the skills and knowledge that they possess at the time of the assessment.

The process of implementing a new model of assessment can be challenging. Many colleges lack the resources to provide a range of comprehensive assessments. The trend toward state standardization of examinations and cutoff scores also poses a challenge to institutions that wish to implement more individualized assessment strategies. While standardization may send more consistent messages about college-ready standards and facilitate cross-state research on student progress, it may work against a more tailored approach, in which colleges might select a range of assessments to place students into different interventions.

Broad reform of assessment and remedial practices may be necessary, but it is unlikely to happen quickly or easily. In the meantime, schools are increasingly using early assessment strategies, offering college placement tests to high school students in order to provide information on skills deficiencies well before college entry. This makes high schools responsible for remediation and may forestall reform of college placement tests and instruction. The California State University system's Early Assessment Program is beginning to yield evidence that participation reduces students' probability of needing 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 math and six percentage points in reading.

While there is more evidence on the need for reform than evidence on what type of reform would work best, this is not cause for discouragement. Some of the alternatives discussed above 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. Institutions could experiment with using placement tests (or multiple measures) for targeting 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. Future research should explore whether current assessment and placement policies have heterogeneous effects—it may be that the current system works well for some subset of students but does not consistently identify who those students are.

While the field has yet to reach a consensus regarding

the best directions for assessment reform, it is clear that change is needed in order to dramatically improve persistence and graduation rates. 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. (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.
- Berger, D. M. (1997). Mandatory assessment and placement: The view from an English department. *New Directions for Community Colleges*, 100, 33–41.
- 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.
- College Board. (2003). *ACCUPLACER OnLine: Technical manual*. New York, NY: College Board.
- College Board. (2007). *ACCUPLACER coordinator's guide*. New York, NY: College Board.
- 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.
- 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.
- 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.
- 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.
- 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 and 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.
- 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.
- Perin, D. (2006). Can community colleges protect both access and standards? The problem of remediation. *Teachers College Record*, 108(3), 339–373.
- Primary Research Group. (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.
- 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.

Funding for this research was provided by the Bill & Melinda Gates Foundation. This Brief is based on CCRC Working Paper No. 19, part of the CCRC Assessment of Evidence Series, which is available for download free of charge at <http://ccrc.tc.columbia.edu>.

Katherine L. Hughes is the Assistant Director for Work and Education Reform Research at the Community College Research Center, Teachers College, Columbia University.

Judith Scott-Clayton is Assistant Professor of Economics and Education at Teachers College, Columbia University, and a Senior Research Associate at the Community College Research Center, Teachers College, Columbia University.