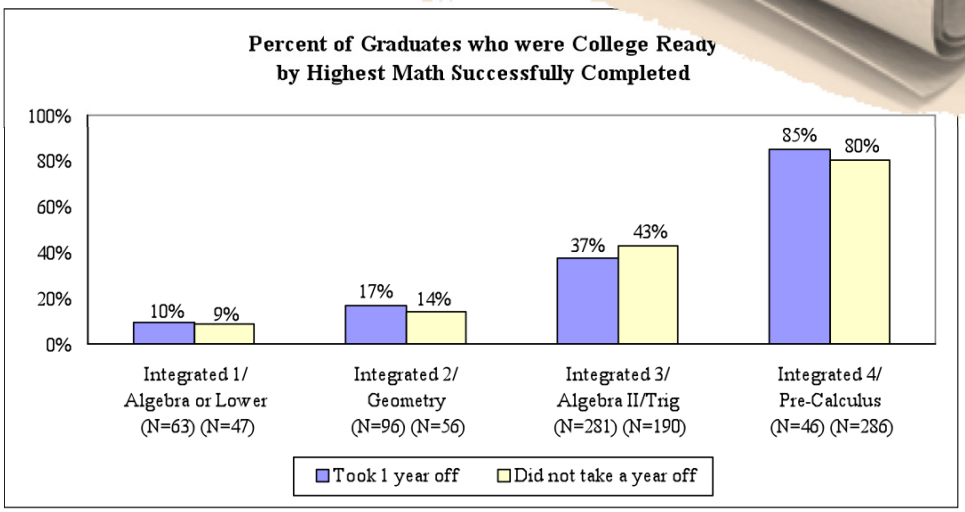
The WERA Educational Journal

Volume 4 Number 1 December 2011



Validity of College and Career Readiness Assessments

By Jack B. Monpas-Huber, Ph.D.

College and career readiness is quickly becoming the primary outcome of public education in the United States. Probably the clearest sign of commitment to this is the current national movement of states to adopt the Common Core State Standards (CCSS) and align with one of the two assessment consortia: the SMARTER Balanced Assessment Consortium (SBAC) or the Partnership for the Assessment of Readiness for College and Careers (PARCC). The purpose of the CCSS in mathematics and English Language Arts is to define what students should know and be able to do to be successful in college and career, and the SBAC and PARCC consortia will develop "next generation" assessments of the new standards. Washington State recently joined this movement by adopting the CCSS and, as a member of SBAC, committing to assess students with the new assessments in 2014-15.

As with any large-scale assessment program, the CCSS and consortium assessments will bring new inferences, decisions, and consequences. Districts and schools will realign their local "power standards" (Ainsworth, 2003), curricular materials, and assessment systems with these new national standards and assessments in order to meet new expectations of performance. Student scores will inform decisions about placement and instructional intervention and broader system-level inferences about curriculum and program effectiveness. Error in the scores will cause at least some students to be misclassified and educators to draw at least some wrong conclusions about instructional effectiveness. To acknowledge these consequences is to consider the validity of assessment, which is "the most fundamental consideration in developing and evaluating tests" (AERA, APA, & NCME, 1999). In this paper, I offer initial thoughts on the validity of college and career readiness assessment. Although there are many possible validity issues to explore (Messick, 1989), here I consider two: operational definitions of college and career readiness and the predictive validity of college and career readiness assessment.

Operational Definitions of College and Career Readiness

The first issue concerns the operational definition college and career readiness: What is college and career readiness, and how is it measured operationally? What is the validity of inferences from operational measurements to the reality of college and career readiness?

Probably the most familiar definitions of college readiness are empirical in nature and come from existing large-scale assessments. The major college testing companies publish "benchmarks" of college readiness based on correlation research into the relationships among exam scores, high school grades, and subsequent grades in college courses. ACT has developed college-ready benchmarks based on the relationship between ACT exam scores and grades in first-year college courses (ACT, 2011), and the College Board now offers similar benchmarks based on the PSAT and SAT (Wyatt et al., 2011).

College readiness is so defined as a level of achieved competency (observed in a test score) associated with a probability of earning particular grades in college level coursework. The contribution of this work is that it imbues test scores with meaning about their predictive value. Students can use the benchmarks to gauge their readiness for college level work, and schools can analyze summary-level benchmark data to evaluate how well their curriculum and instructional program are preparing students for college.

A related line of work involves the 12th Grade National Assessment of Educational Progress (NAEP). As the nation's only large-scale assessment of 12th grade students, the NAEP offers useful information about American high school students' preparedness for college and career level work. The Technical Panel on 12th Grade Preparedness Research recommended a series of validity studies to examine what inferences the 12th Grade NAEP can provide about preparedness for college and career (Loomis, 2011). The Panel recommended content alignment studies between the 12th grade NAEP and other assessments and judgmental standard setting studies to set cut scores on the 12th grade NAEP scale using definitions of preparedness for specific post-secondary activities. It also called for statistical studies linking the 12th grade NAEP to other assessments and a survey study to collect data regarding cut scores on other assessments used for placement decisions. It recommended benchmark studies to collect NAEP data on the students who have entered post-secondary activities (National Assessment Governing Board, 2009). Here again, college readiness is defined as scores on a scale along which certain scores may be defined, either statistically or judgmentally, as meaningful criteria for college or career readiness. An interesting aspect of this work is the implications for different kinds of agreement or disagreement about benchmarks for college and career readiness. Scores defined as college and career ready by one perspective (such as judgmental standard setting methods) may agree or conflict with scores defined as college and career by another perspective (such as benchmark studies of what happens to NAEP examinees once they enter college).

Both the college entrance exam benchmarks and the NAEP work are empirical, defining college and career readiness in terms of scores on scales that derive their significance from sophisticated statistical work based on large sets of quantitative data. The value of an empirical definition is that it facilitates verifiable generalizations about what is generally true across

Continued on next page

large population of people. In this respect it can counterbalance anecdotal definitions of college and career readiness based on individual case studies that tend to appear in the mainstream media.

However, there is more to college and career readiness than test scores. These empirical definitions of college readiness really describe "preparedness." According to NAGB (2009), "Preparedness focuses on academic qualifications, which are measured by NAEP. Readiness includes behavioral aspects of student performance—time management, persistence, and interpersonal skills, for example—which are not measured by NAEP." Conley (2007, 2011a, 2011b) calls attention a broader range of academic behaviors, skills and strategies that are necessary for students to succeed in college but are not necessarily taught in all high schools:

> Recent research has shed light on several other key components of college success. Most relevant for this paper are a range of cognitive and metacognitive capabilities, often referred to as *key cognitive strategies*, which have been consistently and emphatically identified by those who teach entry-level college courses as being of equal or greater importance than any specific content knowledge taught in high school. Examples of key cognitive strategies include analysis, interpretation, precision and accuracy, problem solving, and reasoning (Conley, 2011a, p. 1).

At this point, college and career readiness is primarily defined in empirical terms based on well-established measures. The SBAC and PARCC assessments will add to the current body of measures. However, the reality of college and career readiness is more complex than can be captured in test scores and course grades. As time passes and the gap between established empirical measures and the complexity of college and career readiness comes into clearer focus, it may important to consider a wider range of indicators and data.

Predicting College and Career Readiness

The second issue concerns the validity of inferences about two kinds of predictions: predictions about growth from elementary level measurements toward college and career readiness, and predictions from high school benchmarks about success in college and career. The original charge to the SBAC and PARCC assessment consortia was to measure the Common Core State Standards in English language arts and mathematics in grades 3 through 8 and high school so that all students leave high school prepared for postsecondary success (Duncan, 2010, p. 18175). The SBAC in particular aims to construct assessments that will be based on a vertical scale of achievement that will facilitate inferences about student growth from grade 3 to a high school criterion of college and career readiness (Washington State, 2010). How valid are these growth inferences? To what extent can measurements taken in elementary grades provide meaningful inferences about growth toward a high school criterion of college and career readiness?

Recent research by researchers at the University of Iowa offers a picture of how such a system could work. Furgol, Fina, and Welch (2011) and Dunbar and Welch (2011) asked whether an assessment, such as the Iowa tests, could be used to identify students that are on track for college and career readiness. They looked at longitudinal data from a large sample of Iowa students who took the Iowa Tests of Basic Skills each year from grades 5 to 11 and then took the ACT tests. They found, first, strong correlations between scores from the Iowa Tests at all grade levels and the ACT, which serves as some initial evidence of predictive validity. One key feature of the Iowa Tests is that they use a vertical scale that facilitates estimation and tracking of growth over time. Furgol, Fina, and Welch (2011) explored how to use this scale to determine cut scores on grade level achievement tests prior to 11th grade. They determined "scale scores required at each grade level to achieve the same relative standing within grade as the Iowa scale score corresponding to the ACT college readiness benchmark" (15). The authors concluded that scores from a vertical scale can be used to generate messages at lower grade levels to parents and children about progress toward college and career readiness benchmarks in high school. These studies thus represent initial evidence for the predictive validity of growth inferences about college and career readiness benchmarks.

Another aspect of this criterion-related predictive validity is the extent to which college readiness benchmarks are predictive of success in college. As mentioned above, testing companies have already established targets in the forms of the ACT Benchmarks (ACT, 2010) and the College Board Readiness Index (College Board, 2010) based on the empirical relationship between scores on a college entrance exam scores and success in college (usually first-year undergraduate GPA or grades of B or C in entry-level college courses). This kind of work has played a substantial role in the admissions policies of the nation's colleges and universities, but it is not free of validity threats (Atkinson and Geiser, 2009). The strongest predictor of success in college (as measured by undergraduate GPA) is the high school GPA, followed by the college entrance exam scores, but the best of these models only explain 25% of the variance in undergraduate GPA (Atkinson and Geiser, 2009). How well will the SBAC and PARCC assessments perform as predictors of success in college? Possibly the CCSS will better represent the expectations of competency for most colleges and universities and careers. In addition, the computer adaptive design of the assessments will improve the reliability of the scores, and by extension, their value as predictors of college and career readiness.

Discussion

As with any large-scale assessment program, the CCSS consortium assessments will bring new inferences,

Continued on next page

decisions, and consequences, and these will raise additional validity issues which will prompt the search for the appropriate validity evidence (Messick, 1989). This paper raised two initial questions about the validity of current empirical definitions of college and career readiness, and inferences based on prediction. SBAC and PARCC will add robust new measures to a landscape already populated with indicators, and contribute to a measurement tradition that, arguably, needs to expand to include new indicators of college and career readiness. There also appears to be growing validity evidence for predictions about growth from elementary grades toward high school benchmarks, and the evidence base for predictions about high school benchmarks to postsecondary success will surely grow as these assessments are implemented in the years to come.

References

ACT. (2011). *College readiness standards: For EXPLORE, PLAN, and the ACT*. Retrieved from http://www.act.org/standard/pdf/CRS.pdf

Ainsworth, L. (2003). *Power standards: Identifying the standards that matter most*. Englewood, CO: Advanced Learning Press.

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

Atkinson, R. C. & Geiser, S. (2009). *Reflections on a century of college admissions testing*. *Educational Researcher* 38(9), 665-676.

College Board. (2010). *The development of a multidimensional college readiness index*. College Board Research Report No. 2010-3. Retrieved from http://professionals.collegeboard.com/profdownload/pdf/10b_2084_DevMultiDimenRR_WEB_100618.pdf

Conley, D. T. (2011a). *Redefining college readiness*, Volume 5. Eugene, OR: Educational Policy Improvement Center.

Conley, D. T. (2011b). *Defining and measuring college and career readiness*. Presentation for CCSSO's PNxGL Innovation Labs. Retrieved from http://www.epiconline.org/files/pdf/20110921_CCSSO.pdf

Conley, D. T. (2007). *The challenge of college readiness*. *Educational Leadership*, September, 1-6.

Duncan, A. (2010). *Overview Information; Race to the Top Fund Assessment Program; Notice Inviting Applications for New Awards for Fiscal Year (FY) 2010*. (FR Doc. 2010-8176). *Federal Register*, April 6, 18175-18185. Washington, DC: Government Printing Office.

Furgol, K., Fina, A., & Welch, C. (2011). *Establishing validity evidence to assess college readiness through a vertical scale*. Paper presented at annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Loomis, S. (2011). *A validity framework for reporting preparedness of 12th graders for post-secondary college and career courses*. Paper presented at annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Messick, S. (1989). *Validity*. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103.). New York: American Council on Education and Macmillan.

National Assessment Governing Board. (2009). *Making New Links, 12th Grade and Beyond: Technical Panel on 12th Grade Preparedness Research Final Report*. Retrieved from http://www.nagb.org/publications/PreparednessFinalReport.pdf

Washington State. (2010). *Race to the Top Assessment Program Application for New Grants, Comprehensive Assessment Systems, CFDA Number 84.395B*. Retrieved from http://k12.wa.us/SMARTER/pubdocs/SBAC_Narrative.pdf

Welch, C., & Dunbar, S. (2011). *K-12 assessments and college readiness: Necessary validity evidence for educators, teachers, and parents*. Paper presented at annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Wyatt, J., Kobrin, J., Wiley, A., Camara, W. J., & Proestler, N. (2011). *SAT benchmarks: Development of a college readiness benchmark and its relationship to secondary and postsecondary school performance*. Research Report 2011-5. Retrieved from http://professionals.collegeboard.com/profdownload/pdf/RR2011-5.pdf.

-Jack Monpas-Huber is Director of Assessment and Student Information at Shoreline Public Schools and a Past President of the National Association of Test Directors. Contact him at jack.monpas.huber@shorelineschools.org.