

What Should University Admissions Tests Predict?

Steven E. Stemler

*Department of Psychology
Wesleyan University*

University admissions tests should predict an applicant's ability to succeed in college, but how should this success be defined and measured? The status quo has been to use 1st-year grade point average (FYGPA) as the key indicator of college success, but a review of documents such as university mission statements reveals that universities expect students to develop a broad range of skills that are not always fully captured by FYGPA. In this article, evidence related to college and university documents are reviewed and analyzed for common links with regard to the essential capabilities these institutions purport to seek and to develop in their students. A conceptual model outlining what outcomes admissions tests ought to predict is then presented and discussed. Finally, the article considers whether admissions testing ought to be based on an applicant's aptitude, ability, or achievement in the essential skill areas that most universities aim to develop in their students.

How should colleges decide which applicants they should admit, and what is the proper role of testing in this process? Although there have been many technical critiques throughout the years about the validity of admissions tests (Kobrin, Camara, & Milewski, 2002; Leonard & Jiang, 1999; Noble, 2004; Sternberg & The Rainbow Project Collaborators, 2006; Zwick, 2004), there has been comparatively less discussion about the broader purposes of admissions testing conceptually (Lemann, 2000). Thus, it is worth examining the extent to which tests used in the college admissions process are aligned with the mission and objectives of the institutions they are intended to serve. This article is divided into three sections. In the first section, the various purposes and objectives of higher education are examined based on multiple data sources. The second section of the article addresses the question of what outcomes admissions tests should predict. Finally, in the third section, different approaches to admissions testing are considered and their potential alignment with institutional objectives is discussed.

The theoretical framework guiding this article is drawn from the field of program evaluation (Fitzpatrick, Sanders, & Worthen, 2003; Madaus, Scriven, & Stufflebeam, 1996). Specifically, the current analysis is guided by the objectives-based approach to program evaluation (Tyler, 1990). The

Tyler model stresses the importance of alignment between program objectives, implementation (e.g., curricular and extracurricular activities), and assessment. There has been some discussion in the education literature about the alignment between instructional objectives and high-stakes assessments (Martone & Sireci, 2009; Roach, Niebling, & Kurz, 2008); however, there has been far less discussion about the overarching objectives of higher education and the alignment of these objectives with curricular and cocurricular goals and assessments.

WHAT DO COLLEGES EXIST TO DO?

Colleges and universities serve multiple functions in society. They serve as a mechanism for professional credentialing (Labaree, 1997), as a hub of social and intellectual activity in the local community (Martin, Smith, & Phillips, 2005), and as a platform for innovation and the discovery of new technologies that advance society (e.g., Google), just to name a few. At the most direct level, however, the primary audience for institutions of higher education is their students, and their primary function is student development. Students generally pursue higher education in order to broaden their exposure to ideas that enrich their thinking and to develop skills that will help them to enjoy satisfying careers upon graduation.

Although opinions abound regarding the primary purposes of higher education, the issue can be systematically and

Correspondence should be addressed to Steven E. Stemler, Department of Psychology, Wesleyan University, 207 High Street, Middletown, CT 06459. E-mail: sstemler@wesleyan.edu

empirically studied. One approach to examining the question of purpose is to content analyze school documents, such as their mission statements (Stemler & Bebell, 1999, 2012; Stemler, Bebell, & Son nabend, 2011). Schmitt (2012/this issue) has presented some data from a small-scale exploration of university mission statements undertaken by his research team. In that study, they found that most of the 35 colleges and universities in their sample endorsed 12 themes: (a) knowledge and mastery of general principles, (b) continuous learning and intellectual interest and curiosity, (c) artistic and cultural appreciation, (d) multicultural appreciation, (e) leadership, (f) interpersonal skills, (g) social responsibility, (h) physical and psychological health, (i) career orientation, (j) adaptability, (k) perseverance, and (l) ethics.

In addition to examining university mission statements, a second useful source of data, which is often developed by university faculty and more specifically targeted toward student development, is a statement of "essential capabilities" that universities expect their students to develop. For example, at Wesleyan University (n.d.), the faculty have developed and agreed upon a set of 10 essential capabilities that the university hopes to develop in its students during their time at the institution. These capabilities are (a) writing; (b) speaking; (c) interpretation; (d) quantitative reasoning; (e) logical reasoning; (f) ethical reasoning; (g) citizenship; (h) designing, creating, and realizing; (i) intercultural literacy; and (j) information literacy.

To determine whether these capabilities were typical or aberrant, the websites of the top 125 liberal arts institutions and top 125 national universities of 2010 cited in *U.S. News & World Report* ("Best Colleges," 2010) were searched for documents describing the specific student outcomes each institution aims to foster. The terms that were used to search the school websites included learning goals, expectations, capabilities, core competencies, and educational outcomes. A total of 35% of the liberal arts colleges and 32% of the national universities had these types of documents readily available on their websites.

Table 1 provides some evidence of the extent to which the 10 capabilities of interest to the Wesleyan faculty were of broader interest to other institutions of higher education across the United States that differ by designation, rank, and geographical area. Indeed, it appears that many of these schools share similar aims (for a detailed breakdown, see webtables at <http://www.purposeofschool.com>). It is particularly interesting to note that both types of institutions emphasized writing and intercultural literacy. Furthermore, more than two thirds of the national universities with statements of student outcomes emphasized citizenship, whereas nearly 60% of liberal arts institutions emphasized speaking skills and ethical reasoning.

Although mission statements and statements of essential capabilities or learning outcomes provide one potentially useful source of information regarding the purpose of higher education, it is important to acknowledge that such state-

TABLE 1
Essential Capabilities Endorsed by Higher Education
Institutions

Capability	Liberal Arts Colleges ^a	National Universities ^b
Writing	73%	65%
Speaking	59%	50%
Interpretation	45%	20%
Quantitative reasoning	54%	72%
Logical reasoning	52%	35%
Ethical reasoning	59%	45%
Citizenship	41%	68%
Designing, creating, realizing	43%	35%
Intercultural literacy	68%	85%
Informational literacy	40%	20%

^a $n = 44$. ^b $n = 40$.

ments are not always developed or universally adopted by faculty. Indeed, at present, there is no good empirical data to indicate who typically develops these statements or the extent to which faculty at different institutions embrace them. Consequently, in addition to document analysis, another way to approach the question of purpose is to use survey research methods.

As part of the Liberal Education and America's Promise Initiative, the Association of American Colleges and Universities commissioned Peter D. Hart Research Associates to conduct a series of focus groups as well as a national survey to determine if employers value liberal education. The study (Association of American Colleges and Universities, 2010) revealed that the outcomes of higher education considered important by employers include the ability to communicate effectively, orally and in writing (89% of employers surveyed); critical thinking and analytical reasoning skills (81%); the ability to apply knowledge and skills to real-world settings through internships or other hands-on experiences (79%); the ability to connect choices and actions to ethical decisions (75%); the ability to analyze and solve complex problems (75%); teamwork skills and the ability to collaborate with others in a diverse group setting (71%); the ability to innovate and be creative (70%); the ability to locate, organize, and evaluate information from multiple sources (68%); the ability to work with numbers and understand statistics (63%); an understanding of the role of the United States in the world (57%); an appreciation for cultural diversity in America and other countries (57%); and civic knowledge, civic participation, and community engagement (52%).

Similarly, an evaluation of recent issues of *Recruiting Trends* (Gardner, 2007), a publication based on information supplied by hundreds of companies and organizations concerning the recruitment of recent college graduates, revealed what skills employers were specifically seeking in their recruits. From an analysis of the publication, one can observe the most recent trends of recruitment that are taking hold in the workforce. In 2002–03, ethics and integrity were

considered the most important competencies. The following year, employers expressed their preference for college students to have better developed skills in communication, personal attributes (work ethic, flexibility, initiative and motivation), teamwork, interpersonal skills, and learning (willing to learn continuously new skills & ideas). Finally, the 2005–06 issue observes that an emerging skill is geographic awareness and a global understanding of events as they pertain to the company and industrial sector.

In sum, a variety of different sources of evidence seem to converge upon the major objectives of institutions of higher education, at least from the perspective of student development. It appears that colleges and universities seek to develop expertise in their students in at least two ways. The first is via the development of content area mastery, or what might be called domain-specific knowledge. This particular approach to the development of expertise is in line with that articulated by Ericsson, Krampe, and Tesch-Römer (1993). According to their model, the development of expertise takes many thousands of hours of deliberate practice in a domain. This is manifest in higher education by the conferring of a degree in a particular major area of study (e.g., psychology, astronomy, classics) after the completion of a substantial number of courses.

A second area in which colleges and universities seek to develop expertise is within the context of domain-general abilities. This particular approach to the development of expertise is reflected by Sternberg (1998) with his notion that abilities are forms of developing expertise. Specifically, according to that model, abilities can be developed into competencies, which can then in turn be further developed into expertise. It is worth noting that there has been considerable debate regarding the extent to which cognitive skills can be considered domain general or whether the cognitive skills themselves are situated so intimately within a domain that the skills may not generalize to other domains (e.g., J. R. Anderson, Reder, & Simon, 1996; Brown, Collins, & Duguid, 1989; Lave & Wenger, 1991). Regardless of one's perspective on this matter, there is a general consensus that individuals must have a minimum level of competence/familiarity with a domain of study before they can begin to invoke the kinds of higher order thinking skills that most colleges would like to see developed within their students (e.g., ethical reasoning, intercultural competence, quantitative reasoning). As stated in the Spellings Commission report on the measurement of higher education outcomes, such cognitive skills "should be fostered and developed across the entire educational experience, and in the context of students' major field" (Spellings, 2006, p. 2).

This dual-focus on the development of both domain-specific knowledge and more domain-general cognitive skills is an excellent setup for assessment design. Experts in the field of assessment lean heavily on the development of a table of specifications for identifying how best to measure instructional objectives (Gronlund, 2006). Whereas in the

past these specifications have primarily emphasized domain-specific knowledge (e.g., mastery of particular content knowledge), more recent trends in assessment have seen the development of tables that cross domain-specific knowledge with domain-general cognitive skills (e.g., L. W. Anderson, Krathwohl, Airasian, & Cruikshank, 2001; Gronlund, 2006; Stemler, Grigorenko, Jarvin, & Sternberg, 2006; Stemler, Sternberg, Grigorenko, Jarvin, & Sharpes, 2009; Sternberg, 1997, 1999b).

WHAT SHOULD ADMISSIONS TESTS PREDICT?

Figure 1 presents a conceptual model that will anchor the discussion for the remainder of this article. On the right-hand side of the model are those indicators of success that admissions tests are or should be designed to predict. At a broad level, one set of variables has been classified as indicators of achievement and the other has been classified as indicators of ability. In this section, I discuss each of these in turn, as well as some additional issues to consider when reflecting upon what it is that admissions tests should predict.

Domain-Specific Knowledge and Achievement

For many years now, 1st-year grade point average (FYGPA) has dominated the landscape as the criterion variable that admissions tests are designed to predict. For example, the College Board routinely evaluates its marquee college admissions test (the SAT) by conducting predictive validity studies at different institutions, all of which use FYGPA as the outcome measure (Camara, 2009). Indeed, in the high-stakes world of college admissions testing, any new challenger to the SAT must first demonstrate its capacity to do an equivalent or even better job at predicting the canonical outcome variable of FYGPA. The chief advantage of using FYGPA for comparison purposes is that (nearly) all students enrolled in higher education will have a score on this indicator at the end of their 1st year. Thus, the choice of FYGPA makes good sense from a pragmatic perspective as it, along with enrollment retention, is one of the few indicators on which all students can be directly compared.

Yet FYGPA does possess certain shortcomings as a criterion measure. One of the major shortcomings is the atheoretical nature of the construct. Indeed, one might reasonably ask what exactly FYGPA is supposed to indicate. At some level, FYGPA serves as a proxy for the development of domain-specific knowledge; however, the theoretical rationale for this argument is weaker for FYGPA than for GPA within major, as students are typically enrolled in courses from vastly different content domains during their 1st year. Whatever FYGPA indicates, it is clear that it reflects achievement, rather than

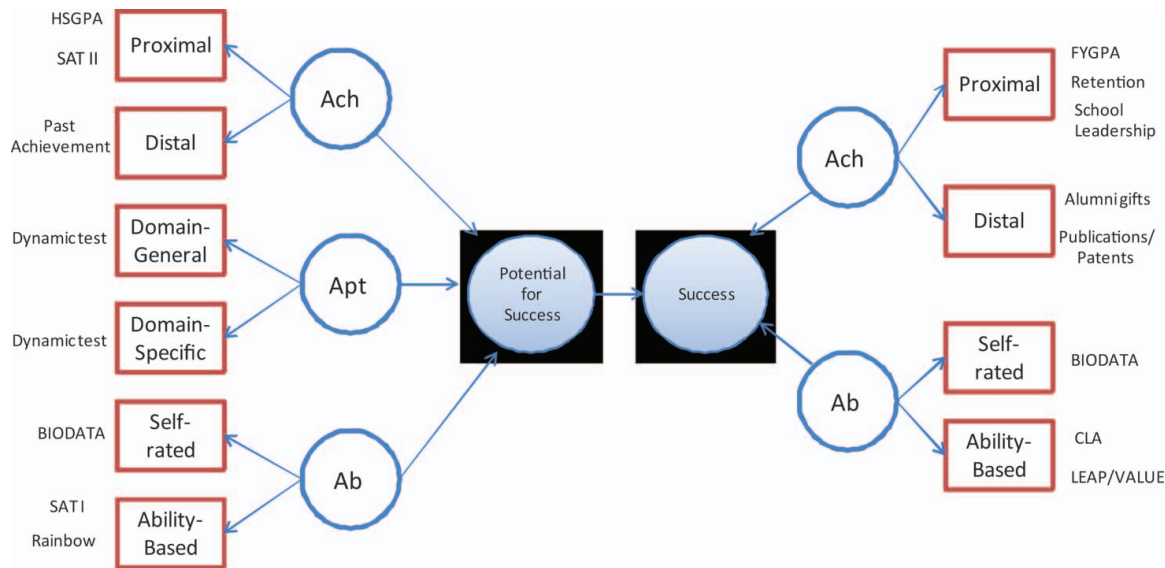


FIGURE 1 Conceptual model. *Note.* HSGPA = high school grade point average; Ach = achievement; Apt = aptitude; Ab = ability; FYGPA = 1st-year grade point average; CLA = Collegiate Learning Assessment (color figure available online).

ability or aptitude, and serves as an indirect measure rather than a direct measure of domain-specific expertise.

Some in the field of assessment (e.g., Airasian & Russell, 2008) have argued that grades can be difficult to interpret because they are so frequently influenced in nonuniform ways by other factors. For example, not everyone goes to the same university. Not everyone in the same university takes the same courses. Not everyone in the same courses has the same instructor. Sometimes the interpersonal relationship (either good or bad) that a student has with an instructor colors the instructor's evaluation of the student's content mastery. Each of these outside factors, and many more, can influence final course grades in ways that are not always related to the cognitive abilities and traits that reside within the student. Nevertheless, given the diverse array of subject matter offerings in higher education and the fact that students will naturally gravitate toward different subject areas, it is perhaps the case that even despite its limitations, FYGPA is the best that we can hope for as a proxy of domain-specific knowledge at this time. Other indicators of achievement that have been used as criterion measures in the past include retention, degree attainment, publications and patents, and alumni giving.

Domain-General Ability

Even if we accept the fact that FYGPA serves as a reasonable proxy for domain-specific knowledge (i.e., content-area mastery), it is clear that FYGPA does not directly reflect the development of domain-general abilities that most universities purportedly value. A prominent 2006 report from the office of the former secretary of education, Margaret Spellings, raised questions about what students are getting for their money from a higher education and what skills or abilities they are

developing that they did not already possess. The commission concluded that "faculty must be at the forefront of defining educational objectives for students and developing meaningful, evidence-based measures of their progress toward those goals" (Spellings, 2006, p. 22). Spurred on by this report, faculty and administrators at higher education institutions across the country have recently begun to grapple with the question of purpose in ways that are amenable to practical measurement (Association of American Colleges & Universities, 2007, 2008). As illustrated previously in this article, most colleges seek to develop a broad range of cognitive skills within their students; however, the field of measurement has not yet settled on the optimal way to measure each of these broader skills. Thus, new measures of these skills that are both domain-specific and domain-general represent an important direction for future research.

There have been some successful preliminary efforts in this regard to date. For example, the Collegiate Learning Assessment (CLA) has taken steps toward this goal with its measurement of critical thinking, analytic reasoning, problem solving, and writing ability (Shavelson, 2007, 2010). Arum and Roksa (2011) recently used the CLA results of students at 24 different 4-year institutions to argue that there is limited learning occurring on college campuses. Their sociological argument as to the reasons behind the poor CLA test results is intriguing and perhaps on point; however, their general critique suffers from an important shortcoming in that it is premised on the erroneous assumption that the CLA captures the full spectrum of skills valued by an institution. Arum and Roksa stated, "Students in general seek to enjoy the benefits of a full collegiate experience that is focused as much on social life as on academic pursuits" and continued by discussing the different incentives for faculty and

administrators, ultimately concluding that “no actors in the system are primarily interested in undergraduate student academic growth, although many are interested in student retention and persistence” (pp. 124–125).

Embedded within this perspective are two erroneous implications: first, that no valuable learning takes place within the context of the students’ social life outside of the classroom, and second, that “academic growth” can be defined solely in terms of the development of critical thinking, problem solving, and writing ability that are assessed by the CLA. Each of these assertions ignores the stated aims of most institutions of higher education (cf. the previous section of this article), which include the development of a much broader range of skills such as intercultural competence, ethical reasoning, information literacy, and quantitative literacy, to name just a few. Although the results from the Arum and Roksa study should give us pause with regard to the success of efforts to develop critical thinking and writing, the CLA does not currently measure the broad spectrum of skills and abilities that institutions themselves purport to seek to develop in their students. The question as to the degree of college student learning raised by Arum and Roksa is important and worthy of further pursuit; however, until new measures of the broader domain-general cognitive skills valued by institutions are developed for use as outcome measures, we are not able to fully address the question of student learning.

Admissions officers would also benefit from the development of broader measures of student learning. According to one admissions officer at an elite institution with whom I spoke while preparing this article, their office rarely tracks in any systematic way whether their admissions procedures “worked.” One potential reason he cited is because there are not very good measures of ability at the outcome level. Thus, it seems that the time is ripe for the development of theoretically based, psychometrically sound indicators of domain-general cognitive abilities that universities would like to see developed in students.

There are several research efforts currently under way in this regard. In addition to the CLA, the Valid Assessment of Learning in Undergraduate Education project of the Liberal Education and America’s Promise initiative of the Association of American Colleges and Universities is another example (<http://www.aacu.org/value/index.cfm>) of an organization that has developed specific rubrics for the development of broader skills and abilities that universities hope to foster within their students. The Measure of Academic Proficiency and Placement developed by Educational Testing Service (ETS) is designed to assess general education skills such as critical thinking, reading, writing, and mathematics for college students (Young, 2007). Other research teams, including the ETS Center for New Constructs (Kyllonen, 2005; Kyllonen, Roberts, & Stankov, 2008) and my own research lab (<http://www.purposeofschool.com>) are also working on developing measures of these broader construct that are relevant for college student populations.

The National Institute for Learning Outcomes Assessment has developed a site to track resources in this area (<http://www.learningoutcomeassessment.org>). Across organizations, there is some consensus on the need to have valid measures of the following cognitive skills: (a) creative thinking, (b) information literacy, (c) civic knowledge and engagement, (d) intercultural competence, (e) ethical reasoning, (f) critical thinking, (g) writing, and (h) quantitative reasoning. Indeed, if we are serious about attempting to improve our power to predict important outcomes of higher education, then we first need to address the criterion problem by developing and integrating broader criterion measures into our predictive validity equations. Yet, whether we are measuring domain-specific achievement or domain-general abilities and their development, there are additional issues to consider when reflecting upon the question of what admissions tests should predict.

Absolute Measures v. Value Added

The Juilliard School and Berklee College of Music are widely considered two of the top postsecondary institutions in the United States for students interested in the performing arts and music. By contrast, MIT, CalTech, and Harvey Mudd are all renowned for their emphases on mathematics, science, and technology. The students who choose to enroll in each of these institutions are both selective and, to some extent, self-selected. Thus, a simple cross-sectional comparison of Harvey Mudd students to Juilliard students with regard to musical ability would likely show Juilliard students achieving at higher levels. By contrast, students attending Harvey Mudd, when compared to students at Juilliard, would probably show higher levels of mathematics achievement.

The question of interest, however, is whether the students in each respective institution are actually enhancing their abilities by attending the institution, or whether they are simply being selected into a context in which they are surrounded with others who share similar talents and abilities. In other words, it is likely that differences in musical and mathematical abilities existed between these two different groups of students before either group took a single course from their respective institution of higher education. Consequently, a cross-sectional comparison of student abilities across different institutions tells us little about the value added by the institution. To address the value-added question, measures that are aligned with the specific objectives of each university must be developed and used and student achievement measured longitudinally (Braun, 2005; Braun, Chudowsky, & Koenig, 2010; McCaffrey, Lockwood, Koretz, & Hamilton, 2003). The extent to which institutions care about gains in such skills over time, as opposed to absolute levels of the skill that their students leave with, should have a profound influence over the approach to admission that is favored.

Proximal v. Distal Indicators

Another important issue to consider with regard to what admissions tests should predict is the tradeoff between proximal and distal/remote indicators (Ruiz-Primo, Shavelson, Hamilton, & Klein, 2002). Naturally, the most interesting outcomes to predict would be those that are most distant in time. Indeed, one of the primary concerns I have heard voiced from faculty colleagues in discussions surrounding the development of higher education outcome measures is the belief that the impact of the lessons learned in and out of the classroom may not fully affect their students until many years after graduation. These faculty members tend to reject the use of proximal indicators such as FYGPA and favor the use of more distal indicators such as indicators of satisfaction with the college experience 10 years postgraduation. The tension, from a psychometric perspective, however, is that the greater the amount of time that elapses between instruction and assessment, the more mediating variables can creep in that impact subsequent performance (for good or ill) making it difficult to link outcomes to predictors.

It seems reasonable to suggest that admissions procedures should be evaluated using both proximal and distal indicators (see Figure 1). From the proximal perspective, admissions officers do sometimes use student retention rates when evaluating the quality of their selection procedures. If students stay at the institutions rather than dropping out or transferring away, then the selection procedure is thought to be working. From the distal perspective, one crude indicator of the effectiveness of selection procedures includes alumni giving to the institution. For example, if a class that is having their 25th reunion donates a large sum to the university, admissions officers may take this as a sign that their admissions procedures worked fairly well, at least for that class of students. Yet the relationship between alumni giving and student learning may be tenuous, at best.

Process v. Product

A final challenge to consider when grappling with the question of what admissions tests should predict is the issue of measuring process or product (Plucker & Renzulli, 1999). Unfortunately, the relationship between processes and products is not always direct. The field of creativity research provides a useful illustration of this tension (Sternberg, 1999a). What determines a creative contribution? Is it the final product (e.g., a painting, sculpture, new mathematical formula) or is a creativity defined by the thinking process used (e.g., divergent and convergent thinking, unusual uses, etc.) to generate a product or solution? The question itself is one that is perennially debated in the literature and one that also interacts with the domain-specific versus domain-general theme underlying many of the arguments in the current article. When viewed through the lens of processes, creativity is typically perceived as a domain-general skill (e.g., Runco, 2009); how-

ever, when viewed through the lens of product or the contribution of eminent individuals, creativity is typically thought to be more domain-specific (Simonton, 2009).

As a practical example of the difficulty of judging a cognitive skill via products or processes, consider the evaluation of George H. W. Bush's controversial decision to raise taxes in the 1990s. As a candidate for president, he famously vowed, "Read my lips, no new taxes." Consequently, when evaluating the product of his decision making (i.e., the decision to raise taxes), one might infer that he had low ethical reasoning because he had contradicted his own promise. However, when viewed from the perspective of the process of decision making, one may recognize that President Bush correctly concluded that the alternatives to not raising taxes would have had a devastating effect on the lives of many individuals. Consequently, when forced with the decision to renege on his campaign promise or do widespread harm via a government shutdown, his choice to raise taxes may, in fact, indicate a high level of ethical reasoning because he did the least harm to the greatest number of people. Politics is replete with examples such as this, in which an evaluation of an individual's ethical reasoning or cultural competence or critical thinking can change depending on whether we view it through the lens of the product or the process.

To determine whether admissions processes are working, it is important to consider what kinds of outcomes admissions tests should predict. Achievement-based outcomes (i.e., products) tend to be what are used in predictive validity equations. Perhaps the most widely used criterion that is used is GPA. Despite the popularity of using FYGPA as an outcome measure in predictive validity equations, many would argue that it is difficult to accurately measure the influence of higher education by taking an achievement-oriented perspective because the lessons students learn from the college experience typically extend far beyond what is taught within the confines of the classroom (Pascarella & Terenzini, 2005).

It is not enough for institutions of higher education to simply look at crude indicators of achievement (e.g., employment status/satisfaction) as evidence that status quo admissions procedures are "working." Rather, what is needed is systematic reflection on the overarching purposes of higher education and the extent to which admissions indicators assess these desired skills accurately (i.e., the cognitive processes). In addition, the development of these skills and abilities should be monitored over time and empirically correlated with subsequent achievement.

WHAT SHOULD ADMISSIONS TESTS MEASURE?

Even in light of some degree of consensus regarding the kinds of expertise that colleges would like to develop in their students, a critical question remains. What is the role of admissions testing in facilitating these objectives? Is the goal to

identify individuals with previously demonstrated levels of competence or accomplishment (i.e., achievement) in areas of value to the institution? Is the goal to identify individuals who exhibit some degree of mastery of a particular skill (i.e., ability) in areas of value to the institution so that these abilities can be further developed? Or is the primary goal to identify individuals with the potential to quickly acquire new skills if placed in the right environmental context (i.e., aptitude)? The answers to these questions have a direct bearing on the approach to selection that one endorses. In this section, the focus will be on the information on the left-hand side of the conceptual model presented in Figure 1.

Aptitude, Ability, and Achievement

Although the terms *aptitude*, *ability*, and *achievement* are, at times, used interchangeably in the literature, they have distinct meanings within the context of measurement that carry with them critical conceptual and philosophical distinctions that are relevant to the present discussion (Stemler & Sternberg, in press).

Aptitude refers to an individual's potential for learning or acquiring a specific skill (Kaplan & Saccuzzo, 2009; Silzer & Church, 2009). It tends to be focused on future potential irrespective of past achievement or present ability. Aptitude also encompasses the rate of speed with which a person will be able to acquire and incorporate new concepts. Aptitude can be viewed as either domain-general (e.g., she is a quick learner) or domain-specific (e.g., he has high mechanical aptitude but low quantitative aptitude).

Ability refers to what an individual is capable of doing at the present moment, often in terms of skills that are measured via speed, accuracy, or both (Kaplan & Saccuzzo, 2009). Abilities indicate a general proficiency for performing tasks and consist of what one has learned from both school and nonschool sources (Airasian & Russell, 2008). The individual with musical ability can perform a piece of music on cue, but we have no way of knowing how that particular ability was developed or whether it is at all predictive of future ability to acquire new information. For one individual, it may take a decade to learn to perform a particular piece of music, and for another individual it may take only a few days. The latter would be said to have higher aptitude than the former; however, both individuals share the same degree of ability in that they demonstrate with equal competence the mastery of a particular skill set.

Achievement refers to demonstrated competence or accomplishment in a particular area. Within the context of testing, achievement tends to be tied to learning that happens in response to explicit instruction (Airasian & Russell, 2008). Achievement also implies some degree of recognition or credentialing. A person with high music ability who has never participated in a concert or been evaluated by a teacher as achieving a particular "grade" level on an instrument lacks

demonstrated achievement, even as she may possess high ability and/or aptitude.

If an individual is a guitar player, we might measure his aptitude for the guitar by teaching him a few new chords or maneuvers and asking him to integrate those things into a piece of music in order to determine how quickly the individual can incorporate new ideas into his repertoire. We might measure his ability by asking him to sight-read a new song or two and evaluate the accuracy with which he is able to play guitar right here and now. We might measure his achievement by examining the number of shows he has performed or his knowledge of chords.

It is important to recognize, however, that although a correlation among these three constructs is often assumed, it is not necessarily the case. It is entirely possible for an individual to demonstrate achievement (e.g., knowledge of music theory) while lacking the ability to play an instrument and showing no aptitude for it. Furthermore, one may demonstrate a particular ability (e.g., strong interpersonal skills) without ever putting those skills to use in any practical manner that would result in the achievement of some desirable goal.

The Assessment of Aptitude

Within the context of college admissions testing, how might one go about assessing aptitude? Although the theoretical concept of aptitude is relatively clear, measuring it directly has proven to be a formidable challenge (Stemler & Sternberg, in press). The SAT was originally conceived of as an attempt to measure scholastic aptitude. Over time, however, it became clear that it was difficult to operationally disentangle what was being measured on this test—aptitude, ability, or achievement, or some combination thereof.

Modern approaches to the assessment of aptitude are most often found in the literature on dynamic assessment (Elliott, 2003; Grigorenko & Sternberg, 1998; Sternberg & Grigorenko, 2001, 2002a). Dynamic assessment is based on Vygotsky's (1978) concept of the zone of proximal development and Feuerstein's concept of mediated learning experiences (Feuerstein & Feuerstein, 1994). According to these authors, it is impossible to understand and assess intellectual potential using a unidirectional, static assessment. Whereas static tests ask a single question in a single way, dynamic tests involve an interaction between the test administrator and the test taker. If a test taker misses a particular item, a graduated series of hints are given to the test taker until the item is answered correctly. Ideally, these hints are scaffolded in such a way as to reflect the types of cognitive errors that the test taker could be making (from the simple, such as not paying attention to directions, to the more complex; see Lidz & Elliott, 2000, for practical examples).

The major limitation to the use of dynamic testing for admissions purposes is the fact that they are necessarily individually administered rather than group-administered tests.

Consequently, the format of the tests does not currently lend itself to the type of large-scale assessment required to be practically useful for the purposes of college admissions. Yet, with the rapid evolution of technology, it is plausible that some adaptive, interactive tests designed to gauge aptitude could be developed within the next decade.

The Assessment of Ability

There are many more options available for assessing ability than there are for assessing aptitude. Recall that abilities represent skills learned both in school and out of school. Thus, someone may exhibit high levels of cultural competence despite never having been explicitly taught about culture within the context of their school curriculum.

There are two major approaches to measurement of abilities that pervade the literature. The first approach involves the self-reporting of abilities and is historically associated with personality-based approaches to assessment (Zeidner, Matthews, & Roberts, 2009). Examples of the self-report approach to assessment include Bar-On's (2000) measure of emotional intelligence and Schmitt and colleagues' biodata measures—specifically the behaviorally anchored rating scales they reported on (Pulakos et al., 2002; Schmitt, 2012/*this issue*). Biodata measures in particular have been shown to predict FYGPA over and above SAT scores and thus show some promise as a supplementary admissions tool. The major challenge to such self-report measures, when they are used for high-stakes selection purposes, is their susceptibility to faking.

A second approach is known as ability-based measurement. This approach to measurement typically requires prompts that are scored against objective criteria for correct and incorrect responses. The assumption of ability-based approaches to admissions testing is that underlying cognitive abilities are the most efficient predictor of who will successfully acquire domain-specific knowledge and/or domain-general abilities. Indeed, a vast amount of research has attempted to link an underlying general cognitive ability (“g”) to student achievement for far more than 100 years (Sternberg & Grigorenko, 2002b). Researchers who espouse this approach are not concerned with the specific sets of abilities that universities wish to develop per se but rather take the position that an underlying domain-general ability (perhaps consisting of working memory capacity) will be the best predictor of whether a particular applicant will be able to acquire the domain-specific knowledge or domain-general abilities of interest to universities. General cognitive ability has been shown to be an extremely robust predictor of a wide variety of work and educational related outcomes (Kuncel, Hezlett, & Ones, 2004; Sackett, Borneman, & Connelly, 2008).

Frey and Detterman (2004) have shown that scores on SAT/ACT tend to correlate very highly with scores on tests of general intellectual ability ($r_s \sim .82-.86$). SAT scores, in turn, tend to be moderately correlate with college GPA

($r = .34-.35$, Bridgeman, McCamley-Jenkins, & Ervin, 2000). Sackett et al. (2008) have found slightly stronger relationship between aggregate SAT scores and FYGPA after controlling for course difficulty and correcting for range restriction and unreliability ($r = .55$, $R^2 = .30$). Nevertheless, although this upper-level estimate represents a fairly good proportion of variance for one instrument to explain on its own, a substantial proportion of the variance in student GPA remains unexplained ($\sim 70\%$ of variance remains unexplained). Thus, some scholars have argued that a single ability indicator alone is not sufficient but that, instead, meaningful incremental variance in FYGPA can be explained by measuring a broader range of cognitive skills.

Recently, Sternberg and colleagues developed tests of broader intellectual abilities, including creative and practical abilities that have been shown to be distinct constructs that add incremental predictive validity to the SAT within the context of college admissions (Sternberg & The Rainbow Project Collaborators, 2006), doubling the predictive power over the SAT alone in a sample of more than 1,000 test takers from 15 different institutions. Other efforts at ability-based measurement within the context of higher education include the situational judgment measures described by Schmitt (2012/*this issue*), a suite of noncognitive ability tests under development at ETS (Kyllonen, 2005; Kyllonen et al., 2008) and the inclusion of social/personality factors (Hannon & McNaughton-Cassill, 2010). The results reported in each of these studies demonstrate that measuring a broader set of abilities can be useful in incrementing in a meaningful way the amount of variance explained in the domain-specific knowledge outcome of FYGPA. However, despite the promise of these new and broader measures of ability, none are yet sufficiently refined for administration at the large-scale level of hundreds of thousands of students, and evaluating further data on their utility at this scale is advisable.

The situation with regard to the left side of Figure 1 is similar to that on the right side of Figure 1 when one considers tests of certain domain-general abilities such as cultural competence and ethical reasoning. As noted in the previous section of this article, the field could benefit from the development and/or adaptation of psychometrically sound measures of these broader constructs that can be integrated into the admissions context. Although both self-report and ability-based measures of some of these constructs do exist (e.g., Bennett, 1993; Caligiuri, Jacobs, & Farr, 2000; Hammer, Bennett, & Wiseman, 2003; Hill & Swanson, 1985; Rest, Cooper, Coder, Masanz, & Anderson, 1974), at present few are sufficiently targeted at the college student population or validated for use in the selection context. There is a reasonable argument to be made for incorporating measures of these abilities into an admissions testing suite of assessments with one of the primary justifications coming from past research that has demonstrated that, “if students are going to pursue a specific course of study, all else equal, assessing knowledge

and skill in that field will yield the most predictive power” (Kuncel & Hezlett, 2010, p. 340). Thus, we can infer that if universities hold as an important goal the development of these broader abilities, then assessments of current levels of ability may be a reasonable predictor of future performance.

The Assessment of Achievement

Most of the assessments related to college admissions decisions are assessments of past achievement. High school GPA is perhaps the most notable indicator of achievement that is used within the admissions contexts. Grades, particularly cumulative grades, do not indicate specific skills but simply indicate that the individual achieved at a particular level of competence according to their instructor’s criteria, whatever that may have been. Although there is no question that high school grades are a strong predictor of future grades in college (Camara, 2009), there is some question as to what construct is being measured by those grades—whether it is motivation, study habits, mastery of the subject matter, ability to determine what the teacher wants, and so on (Airasian & Russell, 2008).

Tests of achievement typically test knowledge (Gronlund, 2006) rather than performance. It is useful at this point to distinguish between knowledge, performance, and achievement. One may possess knowledge but not act on it, and therefore neither perform nor achieve. This can be seen in the case of a low-performing student who lacks motivation and does not bother to respond to test questions despite the fact that he has perfect command of the material (i.e., knowledge).

Performance requires one to carry out a set of procedures that are based on knowledge. One may have explicit knowledge or tacit knowledge of how to execute such procedures or both. Indeed, work by Sternberg and Horvath (1999) has shown that experts are not always aware of the knowledge that they possess that leads them to expert performance; thus, some of the knowledge may be tacit. Simply possessing knowledge is no guarantee of performance, however. A golfer may know exactly where he wants to hit the ball, strategically speaking, but then exhibit poor performance or execution. Thus performance requires knowledge in the first instance but also requires specific skills related to execution or procedure.

Finally, achievement results when one has demonstrated excellent performance within specific environmental constraints (e.g., making the golf shot during a competition; illustrating one’s knowledge within a specific testing session). Performance, which was captured in finite time, yielded an achievement (or product) of some sort that is then judged and evaluated. It may or may not be a valid indicator of one’s knowledge, and it may have been a poor performance, but the achievement is the product that is the end result of the effort within that particular environmental circumstance.

State-mandated exit exams are excellent examples of achievement tests, and it seems reasonable to expect that

performance on such exams could also become a factor that may be of use in college admissions decisions. Some other types of achievements that are often used by admissions officers (Steinberg, 2002) as indicators of the broader constructs of interest (e.g., leadership) are akin to formal positions held (e.g., class president, athletic team captain, number of hours volunteered for an organization). They are, at best, indicators that one has applied one’s abilities toward a tangible product. At worst, however, they serve as poor proxies for ability, as the positions held do not always indicate the skills/abilities of those who hold them. For example, a person may be elected to a leadership position not because of superior critical thinking skills or ethical reasoning but rather simply because there were no other candidates interested in the position.

The Problem of Selection From the Admissions Point of View

Admissions officers face a range of important decisions when considering their options for selection. One question they face is whether it is important that all applicants demonstrate equal, or at least minimum, levels of aptitude, ability, and achievement, or whether some compensatory selection system should be instituted.

Related to this question is the unit of analysis question. Although admissions decisions are often discussed in terms of individual students, the perspective of the admissions officer is more often focused on creating a balanced incoming class at the college or university (Golden, 2006; Steinberg, 2002; Stevens, 2007). In other words, the task of the admissions officer is not simply to evaluate each individual applicant in isolation but to consider the composition of the class as a whole. Thus, the class becomes an entity that itself requires a balance of aptitude, ability, and achievement integrated into it.

The very point of assembling a “class” is predicated on the notion that learning occurs not only inside the classroom but also among peers on the campus in an informal way. Indeed, the skills that many institutions value so highly, such as the development of cultural competence, citizenship, and ethical reasoning, are only partly developed within the context of formal instruction. As large-scale survey research, longitudinal studies, and experimental research have all shown, when students are immersed in a diverse student body, it has positive effects on a wide variety of learning outcomes, retention, satisfaction, leadership skills, and civic engagement. For example, reporting on the results of the National Survey of Student Engagement, which includes data from more than 285,000 nationwide, Kuh (2003) found that students are more likely to be involved in active and collaborative learning when they are exposed to greater diversity, and they are also more likely to report being satisfied with the college experience. From an experimental perspective, Antonio, Chang, Kenny, Levin, and Milem (2003) found that students who were randomly assigned to groups with greater ethnic

diversity showed more complex thinking, as measured by pre- and postdiscussion essays. Further, in an analysis of longitudinal data from a nationally representative sample of 4,200 students, Hurtado (2001) found that studying with peers from diverse backgrounds had a more pronounced impact on self-reported growth in thinking and problem-solving skills than a curriculum emphasizing different perspectives did. Thus, exposure to diverse individuals appears to add a value far greater than simply studying about diversity.

It is worth noting, however, that the vast majority of research on the value of campus diversity for admissions purposes has focused exclusively on the value of ethnic diversity rather than diversity in terms of different cognitive strengths. Imagine, for example, a situation in which students are tested on a much broader array of cognitive skills (e.g., critical thinking, ethical reasoning, intercultural competence) at the point of admission and a class is assembled that is then balanced in terms of individual's demonstrated abilities in different cognitive skill areas. Given the prevalence of the assumption of the value of diversity and its influence over admissions decisions, it would be prudent to empirically evaluate the validity of the claim that creating a "class" facilitates the acquisition of the desired capabilities by measuring their development over time.

DISCUSSION

In this article, I have argued for three main points. The first point is that universities aim to develop two types of expertise in their students: domain-specific knowledge (e.g., via the major, such as physics), and domain-general abilities (e.g., quantitative reasoning, cultural competence). GPA has historically been used as a proxy of domain-specific knowledge and has traditionally been used as the gold standard for investigations into the predictive validity of admissions tests. Yet there is substantial evidence from a variety of diverse data sources indicating that universities and employers also value a specific set of domain-general abilities. At the present time, few psychometrically sound measures of these abilities targeted at the college student population exist. The development of new, more direct measures of a broad range of important skills (e.g., ethical reasoning, cultural competence) that have been articulated as important higher education outcomes is urgently needed. The development of such measures not only would help universities meet the goal of evaluating their value-added to students in cognitive skill areas valued by institutions themselves (Spellings, 2006) but also would help to create more theoretically meaningful criteria for admissions tests to predict as a supplement to the domain-specific knowledge indicated by FYGPA.

Broader tests of abilities could be useful to colleges and universities seeking to understand what students are gaining over time. At the student level, the test results could be used in a formative way to determine what kinds of activities students

might consider pursuing in order to enhance their acquisition of essential capabilities. For example, imagine a sophomore who scores low on an ethical reasoning test at the end of her 1st year. She might decide to enroll in a philosophy course or perhaps consider attending more debates related to ethical issues. It is not difficult to imagine the potential development of computer-automated formative assessments for students, complete with recommendations for enhancing each of the skills in question. Research targeted at scaffolding would demonstrate what kinds of activities might be most beneficial for different types of students.

The adoption of a new set of psychometrically sound measurement instruments designed to measure essential higher education outcomes may open conversations on a broad array of topics. If a university were to find that its students were not scoring well in the area of cultural competence, for example, the results might ultimately influence the kinds of students the university tried to admit and retain (e.g., admit more international students), the kinds of programs the university offers (e.g., broaden the range and emphasis on study abroad programs), and the kinds of activities that are encouraged financially (e.g., cultural programming on campus), as well as emphasis in the classroom (e.g., how findings are interpreted/replicated in different cultures).

The second point is that tests used for the purposes of college admission should be aligned with the stated objectives of the institutions they are intended to serve. Data from a variety of different sources suggest that institutions of higher education attempt to develop students' expertise in specific content domains (e.g., physics, history) while developing in their students cognitive abilities (e.g., critical thinking, quantitative reasoning, ethical reasoning) that may be useful across domains. Yet, despite a broad consensus on the purposes of higher education, tests that are currently used for admissions purposes tend to measure only a fairly narrow range of student abilities.

Although most admissions officers do look at other indicators to get a sense of achievement in broader domains (e.g., positions held as indicators of leadership), such information is currently treated in ways that are highly subjective, as differently admissions officers weight these external factors differently (Steinberg, 2002). Admissions officers would do well to consider incorporating into their decision making more systematic and standardized methods for measuring a broader range of student abilities that are of value to their institutions. In some cases, there already exist tests that can serve as a useful starting point for measuring many of the constructs that institutions of higher education define as essential (e.g., Schmitt et al., 2009; Shavelson, 2007; Sternberg and The Rainbow Project Collaborators, 2006; Zeidner et al., 2009) and many more are currently under development. It is up to admissions officers and administrators to experiment with using these instruments.

The third main point of the article is to highlight the fact that admissions committees tend to rely heavily on measures

of past achievement as a primary mechanism for predicting future performance in college. Although this approach is certainly an excellent starting point, it is not sufficient. Student abilities, particularly in domains of value to institutions of higher education (e.g., ethical reasoning, cultural competence, quantitative reasoning), should also be assessed and taken into account in the admissions process. The problem with measuring only past achievements and not present abilities is that it assumes a correlation between achievement and ability that is not always strong, and it sends a message that product is more important than process. Consider an individual who is elected as a leader of an organization. This particular individual will have “achieved” a degree of leadership; however, the individual may not possess any leadership abilities that are relevant to the job and may be a complete disaster at the helm.

Indeed, it is not too far a stretch to speculate that the incessant focus on product over process has led students to a situation in which they pursue positions or accomplishments to pad their resumes rather than for the opportunities these positions provide to develop their abilities. The emphasis on product over process has fueled a culture that all too often values grades rather than learning and financial gain at the expense of social responsibility. Admissions officers have a role to play in changing this mentality by requiring broader tests of ability in addition to indicators of achievement.

REFERENCES

- Airasian, P. W., & Russell, M. K. (2008). *Classroom assessment* (6th ed.). New York, NY: McGraw Hill.
- Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher*, 25, 5–11.
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., & Cruikshank, K. A. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Boston, MA: Allyn & Bacon.
- Antonio, M. J., Chang, K. H., Kenny, D., Levin, S., & Milem, J. F. (2003). *Effects of racial diversity on complex thinking in college students*. Retrieved from <http://siher.stanford.edu/publications/alphabytitle.html>
- Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: The University of Chicago Press.
- Association of American Colleges and Universities. (2007). *College learning for the new global century: A report from the National Leadership Council for Liberal Education & America's Promise*. Washington, DC: Author.
- Association of American Colleges and Universities. (2008). *Our students best work: A framework for accountability worthy of our mission* (2nd ed.). Washington, DC: Author.
- Association of American Colleges and Universities. (2010). *Raising the bar: Employer's views on college learning in the wake of the economic downturn*. Washington, DC: Hart Research Associates.
- Bar-On, R. (2000). Emotional and social intelligence: Insights from the emotional quotient inventory. In R. Bar-On & J. D. A. Parker (Eds.), *Handbook of emotional intelligence* (pp. 363–388). San Francisco, CA: Jossey-Bass.
- Bennett, J. M. (1993). Toward ethnorelativism: A developmental model of intercultural sensitivity. In R. M. Paige (Ed.), *Education for the intercultural experience* (pp. 21–71). Yarmouth, ME: Intercultural.
- Best colleges. (2010). *U.S. News & World Report*. Retrieved from <http://colleges.usnews.rankingsandreviews.com/best-colleges>
- Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Service. Retrieved from <http://www.ets.org/Media/Research/pdf/PIC-VAM.pdf>
- Braun, H. I., Chudowsky, N., & Koenig, J. (Eds.). (2010). *Getting value out of value-added: Report of a workshop* (Committee on value-added methodology for instructional improvement, program evaluation, and accountability). Washington, DC: National Research Council. Retrieved from <http://www.nap.edu/catalog/12820.html>
- Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade-point average from the revised and recentered SAT I: Reasoning test* (College Board Report No. 2000–1). New York, NY: College Entrance Examination Board.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32–42.
- Caligiuri, P. M., Jacobs, R. R., & Farr, J. L. (2000). The Attitudinal and Behavioral Openness Scale: Scale development and construct validation. *International Journal of Intercultural Relations*, 24, 27–46.
- Camara, W. J. (2009). College admissions testing: Myths and realities in an age of admissions hype. In R. P. Phelps (Eds.), *Correcting fallacies about educational and psychological testing* (pp. 147–180). Washington, DC: American Psychological Association.
- Elliott, J. (2003). Dynamic assessment in educational settings: Realising potential. *Educational Review*, 55(1), 15–32.
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, 363–406.
- Feuerstein, R., & Feuerstein, S. (1994). Mediated learning experience: A theoretical review. In R. Feuerstein, P. S. Klein, & A. J. Tannebaum (Eds.), *Mediated learning experience (MLE): Theoretical, psychosocial and learning implications* (pp. 3–51). London: Freund.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2003). *Program evaluation: Alternative approaches and practical guidelines* (3rd ed.). Boston, MA: Allyn & Bacon.
- Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment of g? *Psychological Science*, 15, 373–377.
- Gardner, P. D. (2007). *Recruiting trends: 2006–2007*. East Lansing: College Employment Research Institute, Michigan State University.
- Golden, D. (2006). *The price of admission: How America's ruling class buys its way into elite colleges—and who gets left outside the gates*. New York, NY: Three Rivers Press.
- Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin*, 124(1), 75–111.
- Gronlund, N. E. (2006). *Assessment of student achievement* (8th ed.). Boston, MA: Allyn & Bacon.
- Hammer, M. R., Bennett, M. J., & Wiseman, R. (2003). Measuring intercultural sensitivity: The intercultural development inventory. *International Journal of Intercultural Relations: Special Training Issue*, 27, 421–443.
- Hannon, B., & McNaughton-Cassill, M. (2010). SAT performance: Understanding the contributions of cognitive/learning and social/personality factors. *Applied Cognitive Psychology*, 25, 528–535. doi:10.1002/acp.1725
- Hill, G., & Swanson, H. L. (1985). Construct validity and reliability of the Ethical Behavior Rating Scale. *Educational and Psychological Measurement*, 45, 285–292.
- Hurtado, S. (2001). Linking diversity and educational purpose: How diversity impacts the classroom environment and student development. In G. Orfield & M. Kurlaender (Eds.), *Diversity challenged: Evidence on the impact of affirmative action* (pp. 187–203). Cambridge, MA: Harvard Civil Rights Project.

- Kaplan, R. M., & Saccuzzo, D. P. (2009). *Psychological testing*. Belmont, CA: Wadsworth.
- Kobrin, J. L., Camara, W. J., & Milewski, G. B. (2002). *The utility of the SAT I and SAT II for admissions decisions in California and the nation* (College Board Report No. 2002-6). New York, NY: College Entrance Examination Board.
- Kuh, G. (2003, March–April). What we're learning about student engagement from NSSE. *Change*, pp. 24–32.
- Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science*, 19, 339–345.
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, 86(1), 148–161.
- Kyllonen, P. C. (2005, September). The case for noncognitive assessments. *ETS R&D Connections*, pp. 1–7.
- Kyllonen, P. C., Roberts, R. D., & Stankov, L. (Eds.). (2008). *Extending intelligence: Enhancements and new constructs*. New York, NY: Erlbaum.
- Labaree, D. F. (1997). *How to succeed in school without really learning*. New Haven, CT: Yale University Press.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Lemann, N. (2000). *The big test: The secret history of the American meritocracy*. New York, NY: Farrar, Straus, and Giroux.
- Leonard, D., & Jiang, J. (1999). Gender bias and the college prediction of the SATs: A cry of despair. *Research in Higher Education*, 40, 375–408.
- Lidz, C., & Elliott, J. G. (Eds.). (2000). *Dynamic assessment: Prevailing models and applications*. New York, NY: JAI.
- Madaus, G. F., Scriven, M., & Stufflebeam, D. (Eds.). (1996). *Evaluation models: Viewpoints on educational and human services evaluation*. Boston, MA: Kluwer-Nijhoff.
- Martin, L. L., Smith, H. P., & Phillips, W. (2005). Bridging 'Town & Gown' through innovative university-community partnerships. *The Innovation Journal*, 10(2), 1–16.
- Martone, A., & Sireci, S. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research*, 79, 1332–1361.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/pubs/research_briefs/RB9050/index1.html
- Noble, J. (2004). The effects of using ACT composite scores and high school averages on college admissions decisions for ethnic groups. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 303–319). New York, NY: Routledge Falmer.
- Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research*. San Francisco, CA: Jossey-Bass.
- Plucker, J. A., & Renzulli, J. S. (1999). Psychometric approaches to the study of human creativity. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 35–61). New York, NY: Cambridge University Press.
- Pulakos, E. D., Schmitt, N., Dorsey, D. W., Arad, S., Hedge, J. W., & Borman, W. C. (2002). Predicting adaptive performance: Further tests of a model of adaptability. *Human Performance*, 15, 299–323.
- Rest, J., Cooper, D., Coder, R., Masanz, J., & Anderson, D. (1974). Judging the important issues in moral dilemmas—An objective test of development. *Developmental Psychology*, 10, 491–501.
- Roach, A. T., Niebling, B. C., & Kurz, A. (2008). Evaluating the alignment among curriculum, instruction, and assessment: Implications and applications for research and practice. *Psychology in the Schools*, 45, 158–176.
- Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching*, 39, 369–393.
- Runco, M. A. (2009). Simplifying theories of creativity and revisiting the criterion problem: A comment on Simonton's (2009) hierarchical model of domain-specific disposition, development, and achievement. *Perspectives on Psychological Science*, 4, 462–465.
- Sackett, P. R., Borneman, J. J., & Connelly, B. S. (2008). High-stakes testing in higher education and employment. *American Psychologist*, 63, 215–227.
- Schmitt, N. (2012/this issue). Development of rationale and measures of noncognitive college student potential. *Educational Psychologist*, 47, 18–29.
- Schmitt, N., Bilington, A. Q., Golubovich, J., Keeney, J., Pleskac, T., Reeder, M., . . . Zorrie, M. (2009). *Report on the first-year follow-up of college applicants at twelve universities*. East Lansing: Michigan State University.
- Shavelson, R. J. (2007, January/February). Assessing student learning responsibly. *Change*, pp. 26–33.
- Shavelson, R. J. (2010). *Measuring college learning responsibly: Accountability in a new era*. Stanford, CA: Stanford University Press.
- Silzer, R., & Church, A. H. (2009). The pearls and perils of identifying potential. *Industrial and Organizational Psychology*, 2, 377–412.
- Simonton, D. K. (2009). Varieties of (scientific) creativity: A hierarchical model of domain-specific disposition, development, and achievement. *Perspectives on Psychological Science*, 4, 441–452.
- Spellings, M. (2006). *A test of leadership: Charting the future of U.S. higher education*. Washington, DC: U.S. Department of Education.
- Steinberg, J. (2002). *The gatekeepers: Inside the admissions process of a premier college*. New York, NY: Penguin.
- Stemler, S. E., & Bebell, D. J. (1999, April). *An empirical approach to understanding and analyzing the mission statements of selected educational institutions*. Paper presented at the New England Educational Research Organization, Portsmouth, NH.
- Stemler, S. E., & Bebell, D. J. (2012). *The school mission statement: Values, goals, and identities in American education*. Larchmont, NY: Eye on Education.
- Stemler, S. E., Bebell, D. J., & Sonnabend, L. A. (2011). Using school mission statements for reflection and research. *Educational Administration Quarterly*, 47, 383–420.
- Stemler, S. E., Grigorenko, E. L., Jarvin, L., & Sternberg, R. J. (2006). Using the theory of successful intelligence as a framework for augmenting AP exams in psychology and statistics. *Contemporary Educational Psychology*, 31, 75–108.
- Stemler, S. E., & Sternberg, R. J. (in press). The assessment of aptitude. In K. F. Geisinger (Ed.), *APA handbook on testing and assessment*. Washington, DC: APA.
- Stemler, S. E., Sternberg, R. J., Grigorenko, E. L., Jarvin, L., & Sharpes, K. (2009). Using the theory of successful intelligence as a framework for developing assessments in AP Physics. *Contemporary Educational Psychology*, 34, 195–209.
- Sternberg, R. J. (1997). *Successful intelligence: How practical and creative intelligence determine success in life*. New York, NY: Plume.
- Sternberg, R. J. (1998). Abilities are forms of developing expertise. *Educational Researcher*, 27(3), 11–20.
- Sternberg, R. J. (Ed.). (1999a). *Handbook of creativity*. New York, NY: Cambridge University Press.
- Sternberg, R. J. (1999b). The theory of successful intelligence. *Review of General Psychology*, 3, 292–316.
- Sternberg, R. J., & Grigorenko, E. L. (2001). All testing is dynamic testing. *Issues in Education*, 7, 137–170.
- Sternberg, R. J., & Grigorenko, E. L. (2002a). *Dynamic testing: The nature and measurement of learning potential*. Cambridge, UK: University of Cambridge.
- Sternberg, R. J., & Grigorenko, E. L. (2002b). *The general factor of intelligence: How general is it?* Mahwah, NJ: Erlbaum.
- Sternberg, R. J., & Horvath, J. A. (1999). *Tacit knowledge in professional practice*. Mahwah, NJ: Erlbaum.

- Sternberg, R. J., & The Rainbow Project Collaborators. (2006). The Rainbow Project: Enhancing the SAT through assessments of analytical, practical, and creative skills. *Intelligence*, 34, 321–350.
- Stevens, M. L. (2007). *Creating a class: College admissions and the education of elites*. Cambridge, MA: Harvard University Press.
- Tyler, R. W. (1990). *Basic principles of curriculum and instruction*. Chicago, IL: University of Chicago Press.
- Vygotsky, L. S. (1978). Mind in society. In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), *L.S. Vygotsky mind in society: The development of higher psychological processes* (pp. 19–119). Cambridge, MA: Harvard University Press.
- Wesleyan University. (n.d.). Essential capabilities. Retrieved January 15, 2011, from <http://www.wesleyan.edu/wesleyanplanning/final/essential.html>
- Young, J. W. (2007). *Validity of the Measure of Academic Proficiency and Progress (MAPP) test: A summary from ETS*. Princeton, NJ: Educational Testing Service.
- Zeidner, M., Matthews, G., & Roberts, R. D. (2009). *What we know about emotional intelligence: How it affects learning, work, relationships, and our mental health*. Cambridge, MA: MIT Press.
- Zwick, R. (Ed.). (2004). *Rethinking the SAT: The future of standardized testing in university admissions*. New York, NY: Routledge Falmer.