CHAPTER 13

THE ASSESSMENT OF APTITUDE

Steven E. Stemler and Robert J. Sternberg

Anyone who has ever been a student, teacher, coach, or parent understands that individuals differ with regard to the speed with which they are capable of learning new information. Some people are "fast learners," others are "slow learners," and most are somewhere in between. Indeed, it is this simple observation that gave rise to efforts to systematically and scientifically measure individual differences in aptitude. Yet, the definition of aptitude itself has proven to be somewhat of a moving target over the years.

The construct of aptitude is often mentioned alongside ability and achievement. Many introductory texts on testing and measurement (e.g., Cohen & Swerdlik, 2005; Gregory, 2007; Kaplan & Saccuzzo, 2009) distinguish among these three terms in roughly the following way: Achievements represent *past* accomplishments or performance, abilities are skills that one can perform right here and now in the *present*, and aptitude reveals an individual's capacity for *future* performance. Although this simple heuristic is relatively useful at a general level, the specific definition of aptitude as well as procedures for assessing it remain hotly contested topics.

Thus, the aim of this chapter is twofold. First, it attempts to describe some of the most controversial elements that serve to define aptitude and highlights areas in the literature where researchers differ. Second, it explores current efforts to assess aptitude, particularly in the context of three domains: (a) college admissions testing, (b) talent identification and personnel selection in employment contexts, and (c) classroom assessment.

WHAT IS APTITUDE?

Disagreements about the definition of aptitude tend to revolve around three central themes. The first of these themes concerns the scope of aptitude. Specifically, is aptitude exclusively a cognitive ability or does it involve noncognitive components as well? The second theme has to do with whether aptitude is something that resides solely within the individual (and which therefore is domain general) or whether it is something that is necessarily the product of a person-by-situation interaction (and therefore is domain specific). Finally, the third major theme is whether aptitude is something that is fixed or modifiable. This section examines the arguments associated with each of these three themes.

Does Aptitude Encompass More Than Just Cognition?

Historical perspectives on aptitude. One of the main concerns of psychologists in the 20th century was the identification of individuals who would be most and least likely to succeed or profit from education. Indeed, the original charge handed down to Alfred Binet from the minister of public instruction in Paris, France, in 1904 was to develop a test of intelligence that would identify children who were not sufficiently profiting from their educational experience so that instructional modifications could be made to help these individuals reach greater levels of success (Birney & Stemler, 2007). The test battery he and his colleague, Theodore Simon, developed ultimately consisted of 30 items measuring everything from simple sensory input and memory

DOI: 10.1037/14049-013
APA Handbook of Testing and Assessment in Psychology: Vol. 3. Testing and Assessment in School Psychology and Education,
K. F. Geisinger (Editor-in-Chief)

to verbal abstractions and social comprehension (Binet, 1916/1905; Gregory, 2007).

As the test traveled across the Atlantic and was imported to the United States by Lewis Terman at Stanford University, an important element of the project was lost. The goal of Terman and his new Stanford–Binet test became the identification of individuals on a linear spectrum of intellectual ability, largely for the purposes of personnel selection rather than remediation or fit to instructional program. It is not coincidental that during roughly the same time period, Charles Spearman (1904) had proposed a general theory (*g* theory) of intelligence, which was based on his observation that levels of performance on many different tests tend to correlate positively—a phenomenon known as positive manifold. Thus, an individual's test scores from a variety of assessments were thought to be related to one another by the concept of an underlying general ability residing within the mind of the test taker. This general ability would manifest itself in an intelligence test score, which could then be used to predict potential for future success.

As a result of the pragmatic predictive success of this approach (Schmidt & Hunter, 1998), the concept of "aptitude became nothing more than the predictions made from conventional ability tests. General aptitude became synonymous with intelligence. Scholastic aptitude became synonymous with verbal and quantitative ability" (Snow, 1992, p. 7). Thus, historically speaking, "the picture of aptitude that most psychologists and educators carried around with them was an entity theory of a fixed, single rank order, general-purpose cognitive trait called intelligence" (Snow, 1992, p. 8; see also Volume 2, Chapter 8, this handbook; Chapter 3, this volume).

Factor analytic studies of aptitude. Formally, Spearman's theory was actually a two-factor theory of intelligence in which *g* was the main focus and explained the majority of observed variance, but a second factor *s* was posited to explain any remaining specific variance. The earliest empirical challenge to this theory came from Thurstone (1938), who approached the analysis of data from a different point of view. Thurstone's approach was to derive factors based on the concept of simple structure, which specified that items would load as highly as possible on only one factor and would have near-zero correlations with all other factors. In doing so, he arrived at a theory of intelligence that specified the existence of seven primary mental abilities: word fluency, verbal comprehension, spatial visualization, number facility, associative memory, reasoning, and perceptual speed (Thurstone, 1938).

Despite the potential incompatibility of the findings from Thurstone and Sperman's two approaches, they found resolution through the specification of a hierarchical factor structure, with a single *g* factor at the top and distinct but correlated subfactors underneath (Brody, 2000). Vernon (1950) simplified the model by proposing a general factor at the top and two broad group factors underneath (v:ed, a verbal-numerical-educational factor, and k:m, a practical-mechanical-spatial-physical factor).

In the 1960s, however, Guilford proposed a nonhierarchical model of intelligence that vastly extended the concept. Specifically, he proposed the existence of three major dimensions (operation, product, and content) on which he could classify any ability test. The operation dimension included five possibilities: cognition (knowing), memory, divergent production (generation of alternatives), convergent production, and evaluation. Each operation could be applied to four different types of content: figural, symbolic, semantic, and behavioral. The application of an operation to a content area could result in one of six products: units, classes, relations, systems, transformations, and implications. Thus, the different combinations could result in a potential of 120 distinct aspects of intelligence. Guilford's model was known as the structure of the intellect (Guilford, 1967), and it represents the most expanded view of intelligence in the field. However, empirical support for the structure of the intellect model has been weak (Brody, 1992).

One of the most important theoretical innovations came from Cattell and Horn's *gf–gc* theory (Horn & Cattell, 1966), which decomposed *g* into two factors: fluid and crystallized ability. Crystallized ability represents an individual's knowledge of information that has been learned, whereas fluid

ability represents an individual's ability to learn. Fluid intelligence is thought to be composed of working memory capacity, processing speed, and inductive reasoning (Kane & Engle, 2002). Fluid ability is very close to the concept of a domain-general aptitude.

Perhaps the most comprehensive contribution within the factor-analytic tradition has come from Carroll's (1993) reanalysis of more than 400 data sets containing cognitive ability test scores. As a result of this massive reanalysis, he arrived at a three-stratum model of intelligence, with a general factor at the top of the hierarchy (stratum III), eight broad factors at stratum II (fluid intelligence, crystallized intelligence, general memory and learning, broad visual perception, broad auditory perception, broad retrieval ability, broad cognitive speediness, and processing speed), and more specific factors at stratum I.

The relation of aptitude to noncognitive factors. Although widely accepted by researchers, the concept of aptitude as a single, cognitively oriented entity was not universally adopted. A second perspective on aptitude, most forcefully advocated by Snow (1977, 1978, 1992) was that the concept of aptitude is not limited to cognitive ability. Rather, he argued, other aspects, such as personality, motivation, and self-concept are also important components of aptitude. Consequently, according to Snow, aptitude consists not only of cognitive factors but also of affective and conative processes. To be clear, within this context, cognitive components refer to analysis and interpretation. Affective components refer to emotions and feelings. Conative components refer to goal setting and will.

Historically speaking, this broader definition of aptitude was consistently advocated by key figures in the field. Both Alfred Binet and David Wechsler, originators of the two most widely used intelligence tests in existence to date, supported this broader view that cognition was indelibly linked to feelings and attitudes (Corno, Cronbach, Kupermintz, & Lohman, 2001).

Thus, the idea that aptitude was composed of cognitive, affective, and conative aspects certainly had theoretical appeal even from the very earliest conceptions of aptitude testing. Researchers such as L. L. Thurstone attempted to develop tests of broader abilities that would yield profile scores. Unfortunately, however, "combining diverse scores into a prediction formula increased the power to predict grade average and other broad indices of success over the predictive power of a full-length 'general' test by only a discouragingly small amount" (Corno et al., 2001, p. 16). Questions remained, however, as to whether this disappointing result was attributable to a flawed theory of aptitude as broader than g or whether the result could be explained by technical limitations of the way the tests were operationalized.

The present era has witnessed a resurgence in interest in the measurement of so-called noncognitive factors for predictive purposes (Kyllonen, Roberts, & Stankov, 2008). Noncognitive factors include such constructs as personality dimensions, time management, self-concept, intercultural sensitivity, and motivation. These constructs are presently being investigated for possible operational use in employment and admissions testing by the Educational Testing Service (Kyllonen, 2005; Kyllonen et al., 2008; see also Chapters 14, 15, and 19, this volume). In addition, Silzer and Church (2009a) recently surveyed more than 100 professionals in organizations and consulting firms who had written on, presented on, or been involved with programs aimed at identifying high-potential employees (i.e., high aptitude) and found substantial overlap among organizations with regard to the key factors used in their approach to identifying potential. These factors included "cognitive skills, personality variables, learning variables, leadership skills, motivational variables, performance records, and other factors" (Silzer & Church, 2009a, p. 391).

Another perspective in the literature that lies somewhere between the g-based perspective on aptitude and the measurement of noncognitive abilities as additional components of aptitude may be found in Sternberg's (1985, 1997, 2005) theory of successful intelligence, which focuses on cognitive abilities but proposes a broader range of cognitive abilities than those measured by g. The theory argues that *successful intelligence* is a person's ability

to achieve his or her goals in life, within his or her sociocultural context, by capitalizing on strengths and correcting or compensating for weaknesses, to adapt to, shape, and select environments through a combination of analytical, creative, and practical skills (Sternberg, 2009).

Successful intelligence therefore conceptualizes cognitive ability in a broader way than g theory does and also acknowledges the important role of what some would call noncognitive factors. A variety of empirical studies have yielded multiple sources of validity evidence (e.g., content, construct, criterion) that support the theory (Stemler, Grigorenko, Jarvin, & Sternberg, 2006; Stemler, Sternberg, Grigorenko, Jarvin, & Sharpes, 2009; Sternberg, Ferrari, Clinkenbeard, & Grigorenko, 1996; Sternberg, Grigorenko, Ferrari, & Clinkenbeard, 1999). In addition, the theory has led to practical test results that have overcome the problems of incremental predictive validity encountered by previous efforts to measure scholastic aptitude in broader ways. More is said about this later in this chapter (see the section College Admissions Testing).

Is Aptitude a Personal Trait or a Person-by-Situation Variable?

Perhaps the most interesting of the three central themes in defining aptitude is the question of whether aptitude is a personal trait that resides within an individual or whether aptitude can be fully understood only by examining the interaction between the person and the situation.

The viewpoint that aptitude is a rather stable trait that resides within an individual carries with it the implication of an ability that is domain general. It is this point of view that gives rise to the idea of the "fast learner" versus the "slow learner." Simply by saying that some people are "fast learners," we are making the assumption that their rate of learning will be relatively stable regardless of what it is they are being asked to learn. Thus, individuals who are fast learners in mathematics will also be fast learners on the athletic field and will be quick to learn how to play musical instruments and so on. Regardless of the content they are learning, we are assuming that there is something inherent in their mind that represents a relatively fixed cognitive ability that allows them to quickly identify new patterns and rules in any domain of interest.

Corno et al. (2001) persuasively argued that the conception of aptitude as a fixed trait of an individual was the result of a misinterpretation of Darwin's theory of natural selection—a theory that had argued that adaptation is the result of a match between organism and environment (i.e., person and situation). Instead, Herbert Spencer, in coining the term *social Darwinism*, mistakenly interpreted Darwin's work and widely disseminated the notion that intelligence (and therefore, aptitude) was a domain-general trait residing within an individual. Thus, the argument proceeded, certain individuals were simply more likely than others to adapt to any set of circumstances they might encounter.

An alternative to the trait perspective is the point of view advocated by researchers who suggest that one cannot understand the concept of aptitude without understanding the context of what is being assessed (Cronbach, 1957; Silzer & Church, 2009a, 2009b; Snow, 1977, 1978). For example, an individual may exhibit a rapid rate of acquisition of knowledge in one domain (e.g., physics) but be a hopelessly slow learner in another domain (e.g., the piano). In such a case, the concept of aptitude only makes sense when discussed relative to the situational context (e.g., an aptitude for physics). This point of view assumes that there is no single cognitive trait inside the minds of individuals that will enable the prediction of rate of learning across all content domains and all time periods. Thus, advocates of the person-by-situation position are aligned with a position known as situated cognition. They believe that there are as many different aptitudes as there are situational contexts and therefore eschew the search for general structures of knowledge that hold across person, situation, and time.

In his classic American Psychological Association (APA) presidential address, Cronbach (1957) outlined the limitations inherent in what he called the two disciplines of psychology (experimental and differential). Specifically, he noted that experimental research on instructional situations often ignores variations in the aptitude of participants, whereas correlational studies of differential psychologists

tend to ignore important situational variation. Thus, he proposed the idea of uniting these two branches of psychology via the study of aptitude–treatment interactions (ATIs). Cronbach stated,

> An aptitude, in this context, is a complex of personal characteristics that accounts for an individual's end state after a particular educational treatment . . . [It] includes whatever promotes . . . survival in a particular educational environment, and it may have as much to do with styles of thought and personality variables as with the abilities covered in conventional tests . . . Such a theory deals with aptitude-treatment interactions." (Cronbach, 1967, pp. 23–24, 30, paragraph altered as cited in Corno et al., 2001, p. 20)

Although the concept of ATIs constituted a new theory of aptitude that challenged the conventional conception of aptitude as an individual trait, empirical support for ATI is still contested. Pashler, McDaniel, Rohrer, and Bjork (2008) have argued that there still is only weak evidence for ATIs; however, Sternberg, Grigorenko, and Zhang (2008a, 2008b) presented an alternative point of view. Some results that support the ATI concept are presented later in this chapter (see the section Classroom Assessment).

Is Aptitude Fixed or Modifiable?

A major debate with regard to the definition of aptitude relates to whether aptitude is something that is stable over time or whether it is modifiable, so that people are able to enhance their aptitude. The traditional psychometric view of aptitude is that it is a relatively fixed trait. According to Silzer and Church (2009a), this perspective is currently widely held by many leaders, managers, and human relations professionals who view the concept of potential as an innate individual capacity (e.g., one has a potential to a certain degree and that degree is not changeable).

An alternative perspective holds that aptitude is malleable. Two noteworthy theorists independently advocating this perspective were the Russian psychologist Lev Vygotsky and the Israeli psychologist Reuven Feuerstein. Vygotsky (1934/1978) introduced the concept of the zone of proximal development, which is the difference between the level of performance that is attainable by an individual on his or her own as compared with the performance that same individual can achieve when aided by someone more knowledgeable or experienced in the domain. The zone of proximal development varies between individuals, such that two people may profit differentially from outside help, and it also varies within individuals, such that the zone may be larger for a given individual in some domains (e.g., music) than in others (e.g., writing; Fabio, 2005). Thus, to Vygotsky, one could not fully understand an individual's potential by looking at scores on a static test of ability. A one-shot test can give us, at best, a snapshot of where an individual currently stands (i.e., his or her ability), but it tells us little about the comparative aptitude (i.e., potential for future performance) of two individuals with the same score. Rather, to assess aptitude more precisely, one must give individuals opportunities to demonstrate how quickly they can grasp new concepts with the aid of a more knowledgeable guide.

In a similar vein, Feuerstein and his colleagues advocated for the importance of what they have called *mediated learning experience* (Feuerstein, Rand, & Hoffman, 1979). Mediated learning experiences are conceptually similar to the zone of proximal development in that they require a more knowledgeable mentor to mediate between a performer and a task by guiding the performer along a scaffolded developmental path toward deeper understanding (Feuerstein, Klein, & Tannenbaum, 1991). What happens during the context of this mediated learning experience is that a qualitative change takes place in the individual's cognitive structure (Birney, 2003).

Silzer and Church (2009a) have proposed somewhat of a compromise position between the concept of aptitude as fixed versus aptitude as modifiable, suggesting that certain components of aptitude represent foundational dimensions, whereas other components represent growth dimensions. According to their proposed structure, foundational dimensions are relatively stable traits that include such components as strategic thinking, dealing with

complexity, and interpersonal skills, whereas growth dimensions include components that individuals can develop and expand, including openness to feedback, risk-taking, and achievement orientation.

Although the debate about whether aptitude is fixed or malleable has not yet reached an empirical conclusion, an important related question has been explored by Dweck (2006). In particular, her research focused on individuals' beliefs—what she calls mind-sets—about whether aptitudes are fixed or modifiable. A rather substantial and growing body of research demonstrates that this question has important practical relevance.

Individuals with a fixed mind-set tend to cling more readily to their first impressions of individuals and to believe that those impressions will accurately predict future behaviors (Erdley & Dweck, 1993). Furthermore, they tend to avoid or ignore subsequent information they receive that contradicts their initial impressions, creating a sort of self-fulfilling prophesy in terms of their beliefs about individuals (Gervey, Chiu, Hong, & Dweck, 1999; Plaks, Stroessner, Dweck, & Sherman, 2001). A fixed mind-set is particularly problematic when one is attempting to assess aptitude or potential because, as Heslin, Latham, and VanderWalle (2005) found, managers with fixed mind-sets are more likely to miss potential or even to misidentify those with low potential (i.e., false alarms) based on the rigidity of their initial impressions and their reluctance to take into account additional performance information.

The different points of view with regard to the three questions posed in this section lead directly to different approaches to the assessment of aptitude, as discussed in further detail in the next section of this chapter.

HOW IS APTITUDE ASSESSED?

Although many fields and professions have contributed to the definition and assessment of aptitude, three domains in particular have demonstrated a persistent concern with the construct. These domains are (a) college admissions testing; (b) employment testing, particularly in the context of talent identification and personnel selection; and (c) classroom assessment, particularly as it relates to ATIs and dynamic-assessment techniques. In this second section of the chapter, we illustrate the way in which the answers to the three key questions posed in the first section of this chapter have practical consequences for the assessment of aptitude.

College Admissions Testing

Philosophical paradigms. As Lemann (2000) has pointed out in his excellent history of the SAT, in the early 20th century a large-scale college admissions test could be based on basically four distinct paradigms. The first paradigm was associated with the philosophy of progressive education advocated by John Dewey (1916, 1938). The goal of individuals in this camp was to develop liberal-minded, free-thinking, and tolerant thinkers. They believed that the best route by which to accomplish this goal was to let schools set their own curricula. Thus, from an admissions perspective, what would be required is a test to determine which students across all of the schools had best developed a broad range of important intellectual abilities.

By contrast, the second paradigm came from individuals such as Ben Wood, who were concerned with a strict, standards-based approach to education. Advocates of this position wanted a standardized curriculum across all schools and felt that admissions tests ought to be based primarily on student achievement, which was of course to be aligned with the curriculum. The descendants of this philosophy are making a strong resurgence in the present day with the current push by the federal government for states to adopt "common core standards" on which all students may be tested and compared (U.S. Department of Education, 2010).

The third paradigm was found in the philosophy of educational expansionists, such as George Zook, who believed that the proper role of testing was to identify students in need of remediation. From this perspective, the goal was education for all and the ultimate goal of testing should be to determine the best fit between an individual and the kind of education that will allow that individual to progress up the developmental ladder. This conception aligns

well with the notion of aptitude as a situation-specific and modifiable trait.

The final paradigm, and the one that eventually won out, was drawn from those individuals who were believers in intelligence testing. The goal of individuals in this camp, which included Educational Testing Service founder Henry Chauncey, was to identify students who would be best able to profit from higher education by selecting those with the highest set of test scores on a new, scholastically oriented, intelligence-type test. The test that was developed based on this philosophy originally was known by its full title as the Scholastic Aptitude Test. The model of aptitude on which the SAT was based was that aptitude consists of only a narrow range of cognitive abilities (specifically verbal and quantitative reasoning), that it was a domain-general trait residing within a person, and that it was relatively fixed. Furthermore, the test was atheoretical and concerned primarily with its power to predict college grades.

Critics of the static testing procedures employed in this context were quick to point out that these tests emphasize previously acquired knowledge (e.g., vocabulary) and do not typically assess how an individual responds to changing circumstances or modifications to the test aimed at increasing levels of performance (Brown & French, 1979; Carlson & Wiedl, 1992; Fabio, 2005; Feuerstein et al., 1979). After some years of debate about whether the SAT could best be described as a measure of aptitude, ability, achievement, some combination thereof, or a subset thereof, the test developers have abandoned the concept of using the SAT as an acronym for its larger descriptive title (i.e., Scholastic Aptitude Test) and have instead settled on simply calling the test the SAT.

What is interesting to consider, however, is what a related type of test would look like that is based on a different conception of aptitude. For example, consider a test that is theoretically based, that conceives of aptitude as encompassing a broader set of cognitive or noncognitive skills beyond those measured in a *g*-type assessment, that assumes aptitude involves an interaction between the person and the environment, and that conceives of and measures aptitude as a modifiable quantity, rather than as a fixed entity.

Sternberg and colleagues (Sternberg & the Rainbow Project Collaborators, 2006) developed a novel assessment containing at least some of those features as a supplement to the standard SAT. Their aim was to develop an aptitude test for college admissions that was theoretically based and that measured a broader range of skills than is currently assessed by the SAT. It further would allow individuals to capitalize on their strengths and compensate for their weaknesses within the context of the test. Their test was developed as part of what was known as the Rainbow Project.

The Rainbow Project. The goal of the Rainbow Project was not to replace the SAT but rather to devise tests that would supplement the SAT, measuring cognitive skills that the SAT does not measure, as outlined by Sternberg's theory of successful intelligence (Sternberg, 1997). In addition to multiple-choice tests, the test used three additional measures of creative skills and three additional measures of practical skills.

Creative skills were measured by using a cartoon captioning task, written stories, and oral stories. On the cartoon task, participants were given five cartoons purchased from the archives of the *New Yorker* with the captions removed. The participant's task was to choose three cartoons and to provide a caption for each cartoon. Two trained judges rated all the cartoon captions for cleverness, humor, and originality. A combined creativity score was formed by summing the individual ratings on each dimension. Next, participants were asked to write two stories, spending about 15 min on each, choosing from the following titles: "A Fifth Chance," "2983," "Beyond the Edge," "The Octopus's Sneakers," "It's Moving Backwards," and "Not Enough Time."

A team of four judges was trained to rate the stories for originality, complexity, emotional evocativeness, and descriptiveness. Finally, participants were presented with five sheets of paper, each containing a set of pictures linked by a common theme. For example, participants might receive a sheet of paper with images of a musical theme, a money theme, or a travel theme. Each participant then chose one of the pages and was given 15 min to

formulate a short story and dictate it into a cassette recorder. The dictation period was not to be more than 5 min long. The process was then repeated with another sheet of images so that each participant dictated a total of two oral stories. Six judges were trained to rate the stories for originality, complexity, emotional evocativeness, and descriptiveness.

Practical skills were measured by using three different types of situational–judgment tests. On the first test, participants were shown a series of seven video-based vignettes designed to capture problems encountered in general, everyday life, such as determining what to do when one is asked to write a letter of recommendation for someone one does not know particularly well. On the second test, participants were given a written description of 15 vignettes designed to capture problems encountered in general business-related situations, such as managing tedious tasks or handling a competitive work situation. On the third test, a written inventory presented participants with 15 vignettes that captured problems encountered in general college-related situations, such as handling trips to the bursar's office or dealing with a difficult roommate. In all cases, the vignettes were followed by a variety of different options for how to handle the situation, and participants were asked to rate the quality of each potential response. Participant responses were then scored based on their distance from the group consensus as to the quality of each response.

A total of 1,015 students at 15 different institutions (13 colleges and two high schools) were tested with this new measure. The results showed that these tests significantly and substantially improved on the validity of the SAT for predicting 1st-year college grades (Sternberg & the Rainbow Project Collaborators, 2006), doubling prediction over the SAT alone, and increasing prediction by 50% over SAT and high school grade point average. The test also improved equity: Using the test to admit a class would result in greater ethnic diversity than would using just the SAT or just the SAT and grade point average. In addition, differences in achievement between White students and African American students were reduced on measures of creative skills, and differences in achievement between White and Latino students were greatly reduced on assessments that emphasized practical skills and creative skills.

One of the main contributions of the Rainbow Project is that it demonstrates that universities potentially can do a better job of predicting who is likely to succeed in college (i.e., who has "more" scholastic aptitude) when a broader range of skills are systematically assessed. Furthermore, universities could be in a better position to select an optimal mix of students of diverse skills, which can be particularly beneficial for ethnic-minority students who tend to perform better at these broader skills that traditionally have been undervalued in terms of assessment but that are highly valued in the university and workforce settings.

Further efforts to develop college-admissions tests that capture a broad range of cognitive and noncognitive skills are being vigorously pursued by several different research groups (Kyllonen et al., 2008; Schmitt, Oswald, & Gillespie, 2005; Schmitt et al., 2007; Stemler, 2012; Sternberg, Bonney, Gabora, Karelitz, & Coffin, 2010).

Employment Testing

A second area in which the assessment of aptitude has had a strong historical connection has been within the area of employment testing and personnel selection. There are two areas within this domain that are of particular interest. The first is within the context of the military and the second is within the context of private sector organizations.

Military testing. Some of the earliest systematic and large-scale standardized efforts to measure aptitude in the United States emerged within the context of the military. Although the relationship between large-scale standardized testing for military recruitment purposes and the development of the SAT is widely recognized (Gregory, 2007; Kaplan & Saccuzzo, 2009; Lemann, 2000), what is less commonly recognized is where those two programs diverge. One of the ways they began to diverge was with regard to the fundamental definition of aptitude that each adopted.

During World War II, a collection of aptitude tests was developed to select among men in the Army who applied for pilot training. Toward the

end of the war, DuBois (1947) evaluated test data from a 2-year period. What he found was that the aptitude test battery was highly predictive of graduation from pilot training school, with percentage passing correlating highly with a specific test score. Furthermore, he found that adding reading and mathematics tests to the composite failed to improve correlation with graduation from pilot training (Corno et al., 2001).

Thus, the concept of aptitude as a domain-specific construct that requires attention to the fit between the person and the demands of his or her occupation came to dominate the military definition of aptitude. Indeed, the Armed Services Vocational Aptitude Battery (ASVAB) is now touted by the military as "the most widely used multiple-aptitude test battery in the world" (Today's Military, 2012). Underlying this test is a different conception of aptitude than the large-scale tests historically used for college-admissions purposes, which tend to consider aptitude as a domain-general capacity of an individual.

The ASVAB includes 10 subtests. These subtests include General Science (GS), Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), Numerical Operations (NO), Coding Speed (CS), Auto and Shop Information (AS), Mathematics Knowledge (MK), Mechanical Comprehension (MC), Electronics Information (EI), and Sum of Word Knowledge and Paragraph Comprehension (VE). Scores from each of these subtests are combined in unique ways to determine the potential of the applicant to succeed in a variety of job positions within each of the branches of the military. The Army creates different subscale scores (line scores) for different professions. For example, positions classified as Clerical require a combination of three tests (VE + AR + MK). Combat positions require a minimum score for the combination of AR + CS + AS + MC. Electronics positions require a minimum score for GS + AR + MK + EI, and so on. More specifically, different positions have different required cut scores. For example, a Special Forces Weapons Sergeant requires a minimum score of 110 on the General Technical subscale (VE + AR) and 100 on the Combat subscale.

Recent research with regard to personnel selection among higher level officers in the military has focused on the development of tests that measure a broad range of cognitive and noncognitive factors as well, including capacities such as adaptability and mental flexibility (Matthew, Beckmann, & Sternberg, 2008; Matthew & Stemler, 2008; Pulakos et al., 2002; Mueller-Hanson, Swartout, Hilton, & Nelson, 2009; Stemler, 2009). Thus, just as the trend in college admissions testing has been to expand the conception of aptitude, it appears that so too in the military context, the definition of aptitude as inclusive of both cognitive and noncognitive factors may be gaining momentum.

Talent identification in the private sector. One interesting question to consider is the way in which the answer to the question of "potential for what?" changes when one views the organization, rather than the individual, as the relevant unit of analysis. As Yost and Chang (2009) pointed out, human resource professionals often think of assessing aptitude within the context of a hierarchical organizational structure in which the aim is to identify potential along a particular dimension (e.g., leadership potential). When one steps back and looks at it from an organizational perspective, however, the main purpose of talent identification is to identify unique profiles of strengths and weaknesses among employees to fulfill a variety of different job roles in an organization. Thus, the aim of the organization is to increase the overall potential of the system.

Indeed, as Yost and Chang (2009) wrote,

> In today's business landscape, the environments in which organizations operate are often so dynamic that investing in only a few people for a limited set of roles is risky. As the last two decades have shown, organizational strategies can change quickly and dramatically, requiring a completely different talent mix to meet future challenges. In dynamic markets, organizations can't always anticipate the challenges they will face and the people they will need in order to compete. (p. 442)

Thus, rather than simply using aptitude assessment to identify leadership potential within an organization, a more comprehensive approach to

aptitude assessment involves identifying and capitalizing on the unique blend of strengths and weaknesses of the organization's employees.

Mone, Acritani, and Eisinger (2009) have cautioned, however, that managers sometimes confuse the assessment of current skills that foster immediate promotability with the kinds of traits that are important for long-term potential in a future role. They suggested separating out these two assessments: "More specifically, stating that exclusively looking at current performance over time does not predict success in advanced roles has helped managers broaden their scope and more accurately assess potential" (p. 427).

In their recent focal article in the journal *Industrial and Organizational Psychology*, Silzer and Church (2009a) illustrated that many organizations take the perspective that the search for potential and aptitude involves the observation of a person-by-situation interaction. These authors suggested that, in practice, employees may need to be given a range of bosses and jobs for their underlying potential to be fully assessed. Doing so will help to address questions such as how quickly an employee can adapt to a particular set of demands, a particular kind of task, or a particular type of supervisor.

Certainly one does not have to dig deeply into the literature on social psychology to understand the profound influence the situation can have on the performance of an individual. In a classic study of the Pygmalion effect, Rosenthal and Jacobson (1968/1992) randomly assigned students to one of two classrooms. The students showed similar levels of achievement before instruction; however, the teacher in one classroom was told that the students had been identified by the assessments as "gifted," whereas the teacher in the control classroom had not been told anything. Stunningly, the results showed that upon posttest, the students in the so-called gifted classroom actually significantly outperformed students in the control classroom. The differences were largely attributed to the way in which teachers interacted with students whom they thought had been identified as high potential.

Although the Rosenthal and Jacobson (1968/1992) study has been criticized on methodological grounds, the general findings have been replicated by researchers employing more rigorous techniques (Weinstein, 2002). For example, a study by Eden and Shani (1982) replicated this finding in the business world and found that trainees who were labeled as having great potential for high performance before a simulated training session outperformed those who were not labeled as high potential. As Heslin (2009) has pointed out, however, it is still not entirely clear how this mechanism works. Studies such as those by Silzer and Church (2009a) do not elaborate on the kinds of bosses that are likely to recognize or to overlook their employee's potential. Nevertheless, these studies and others do demonstrate the important influence of situational characteristics on performance. Therefore, it is perhaps not surprising that many individuals in the private sector appear committed to the notion that aptitude involves a person-by-situation interaction.

Classroom Assessment

A third domain in which the assessment of aptitude has been of keen interest has been within the area of classroom assessment. Two lines of research that are particularly relevant to the discussion of aptitude within this domain are ATI studies and dynamic-assessment research.

ATIs. According to the theory of successful intelligence (Sternberg, 1997), different students have different combinations of cognitive skills (e.g., analytic, creative, and practical). Furthermore, the theory is based on the notion that students learn in different ways—that they have different styles of learning (Sternberg, Grigorenko, & Zhang, 2008a, 2008b), just as teachers have different styles of teaching (Spear & Sternberg, 1987).

Teaching for analytical thinking means encouraging students to (a) analyze, (b) critique, (c) judge, (d) compare and contrast, (e) evaluate, and (f) assess. When teachers refer to teaching for "critical thinking," some of them may mean teaching for analytical thinking. An example of an exercise developing such skills would be to ask students to

compare and contrast two works of literature, to evaluate the conclusions drawn from a scientific experiment, or to critique a work of art.

Teaching for creative thinking means encouraging students to (a) create, (b) invent, (c) discover, (d) imagine if . . . , (e) suppose that . . . , and (f) predict. Teaching for creative thinking requires teachers not only to support and encourage creativity but also to act as a role model and to reward creativity when it is displayed (Sternberg & Lubart, 1995; Sternberg & Williams, 1996). Examples of teaching activities might include asking students to design a psychological experiment to test an hypothesis, to invent an alternative ending for a story they have read, or to create a mathematics problem.

Teaching for practical thinking means encouraging students to (a) apply, (b) use, (c) put into practice, (d) implement, (e) employ, and (f) render practical what they know. Such teaching must relate to the real practical needs of the students, not just to what would be practical for individuals other than the students (Sternberg et al., 2000). Examples might include asking students to apply what they have read in a story to their life, use their knowledge of mathematics to balance a checkbook, or persuade someone that an argument they are employing is sound.

To validate the relevance of the theory of successful intelligence in the classroom, researchers have carried out a number of instructional studies with different age-groups and subject matters. (Other kinds of research support are summarized in Sternberg, 1985, 1997, 2003b.)

In one study (Sternberg et al., 1999), the investigators used the Sternberg Triarchic Abilities Test (Sternberg, 2003a), which assesses analytical, creative, and practical skills through multiple-choice and essay items. The test was administered to 326 children across the United States and in some other countries who were identified by their schools as gifted by any standard whatsoever. Children were selected for a summer program in (college-level) psychology if they fell into one of five ability groupings: high analytical, high creative, high practical, high balanced (high in all three abilities), or low balanced (low in all three abilities).

The high school students who came to Yale were then divided into four instructional groups. Students in all four instructional groups used the same introductory psychology textbook (a preliminary version of Sternberg, 1995) and listened to the same psychology lectures. What differed among them was the type of afternoon discussion section to which they were assigned. They were assigned to an instructional condition that emphasized either memory, analytical, creative, or practical instruction. For example, in the memory condition, they might be asked to describe the main tenets of a major theory of depression. In the analytical condition, they might be asked to compare and contrast two theories of depression. In the creative condition, they might be asked to formulate their own theory of depression. In the practical condition, they might be asked how they could use what they had learned about depression to help a friend who was depressed.

Students in all four instructional conditions were evaluated in terms of their performance on homework, a midterm exam, a final exam, and an independent project. Each type of work was evaluated for memory, analytical, creative, and practical quality. Thus, all students were evaluated in exactly the same way.

The results showed that there was an aptitude-treatment interaction whereby students who were placed in instructional conditions that better matched their pattern of abilities outperformed students who were mismatched. In other words, when students are taught at least some of the time in a way that fits how they think, they do better in school. These results suggest that the negative Cronbach and Snow (1977) results for ATIs may have been due to lack of theoretical basis for instruction or theoretical match between instruction and assessment.

Dynamic assessment. Although the concept of dynamic assessment predates the concept of ATI, it has only been much more recently that scholars have attempted to empirically evaluate dynamic assessment procedures as a method for assessing aptitude (Grigorenko & Sternberg, 1998; Haywood & Lidz, 2007; Lidz & Elliott, 2000;

Sternberg & Grigorenko, 2002; see also Chapter 7, this volume).

The basic premise of dynamic assessment is that one cannot truly understand or assess the aptitude of an individual simply by administering a test at one time point and interpreting that test score. Rather, dynamic assessment is based on the theoretical work of Vygotsky (1934/1978) and Feuerstein and Feuerstein (1994), mentioned earlier, both of whom noted the importance of assessing an individual at more than one point in time and comparing the performance of individuals when they are alone with the performance of those same individuals when they are guided by more knowledgeable others. The chief interest of psychologists and educators engaging in dynamic assessment is not where the test takers are now, given their previous educational experience, but where they can be tomorrow, assuming that they are given adequate educational intervention from now on (Grigorenko, 2009).

As Elliott (2003) has noted, *dynamic assessment* is an "umbrella term used to describe a heterogeneous range of approaches" (p. 16). Some advocates of dynamic assessment conceive of aptitude as domain general (Feuerstein et al., 1979), whereas others believe that aptitude is domain specific (Camilleri, 2005; Guthke, 1992). In practice, there are four main approaches to dynamic assessment (Jeltova et al., 2007).

The first approach is referred to as the test–teach–retest approach. This approach is not so far removed from the procedures invoked by many classroom teachers in the 21st century. The difference is often in the level of detail that occurs at the instructional level. Strictly speaking, the test–teach–retest approach, associated primarily with the work of Budoff (1987), involves protocols for pointing out errors that can be developed that are standardized and even automated.

A second approach to dynamic testing has been dubbed the learning-test approach (Beckmann & Guthke, 1995, 1999; Guthke, 1992). Under this model, the participants are given a pre- and posttest with an intervention in between; however, the procedure extends the previous approach by offering a sequential construction of what information and skills are needed to ensure a successful solution. Furthermore, qualitative analyses of errors are used to diagnose learning processes. Although Guthke, Beckmann, and Dobat (1997) found the results from the learning test to be better predictors of knowledge acquisition and knowledge application in the context of complex performance, Hessles and colleagues (Hessles & Hamers, 1993; Hamers, Hessles, & Pennings, 1996) found no increase in the predictive power of learning tests over traditional tests in their sample of test takers.

A third approach to dynamic testing is called the graduated-prompt approach (Campione & Brown, 1987). The idea behind this model is to give the participant a pretest, a hinted stage, a posttest, and a hint-assisted posttest. This procedure has been shown to be predictive of school readiness for students who are ready to respond to intervention in language production (Olswang & Bain, 1996). As Jeltova et al. (2007) pointed out, however, this approach has a few technical problems. One main criticism is that hints differ in helpfulness across students, so there is some difficulty inherent in trying to standardize the utility of different types of hints.

Finally, the fourth main approach to dynamic testing is called testing the limits (Carlson & Wiedl, 1992). The key objective under this model is to find the best match between the individual and the test situation that will evoke the best possible performance. Thus, this approach is a highly person-by-situation-oriented perspective on the concept of aptitude.

Each of the main approaches to dynamic assessment described here share in common the belief that aptitude cannot be assessed as a fixed, latent trait of an individual that is revealed within the context of a one-shot static test. In general, researchers in this area tend to conceptualize aptitude as largely a cognitive capacity that is situation specific and that is malleable.

DISCUSSION

This chapter has outlined three major debates regarding the definition of aptitude. The first of these asks whether aptitude consists of only cognitive elements or whether it also includes noncognitive elements. Debates around this issue have been

played out largely within the domain of college admissions testing. Although there remains some disagreement about this matter, many major theorists in the field in the 21st century suggest that aptitude includes cognition as well as other noncognitive components. Where theorists tend to diverge is largely with regard to how many other components they include under this umbrella.

The second question, which relates to whether aptitude itself is a rather domain-general trait of individuals or whether it is a domain-specific product of a person-by-situation interaction, has largely been emphasized within the field of employment testing. Disagreements on this question can be viewed with respect to a lock-and-key metaphor of aptitude. Theorists advocating a person-by-situation perspective of aptitude view situational characteristics as a lock and personal attributes as the key that opens the door. By contrast, theorists who view aptitude as a domain-general construct that can be found within an individual without much regard to situational factors view personal aptitude as a skeleton key that will open any situational door.

The third major distinction with regard to aptitude is whether it is fixed or modifiable. Recently, the question as to whether aptitude is fixed or malleable has been most heavily emphasized within the context of classroom assessment. Although some traditional theories have tended to conceive of aptitude as a fixed entity, new advances in technology are beginning to enable assessments that allow for more dynamic testing of individuals and that will open new avenues to those who view aptitude as modifiable and something that is best assessed dynamically. Furthermore, recent research by Dweck (2006) and her colleagues has demonstrated the powerful influence that even the mind-set one has with regard to the question of modifiability of aptitude can have profound consequences across a variety of situational contexts, including teaching, employment, and sports.

Different points of view on these fundamental debates in the definition of aptitude will lead to (and have led to) very different approaches to the assessment of aptitude. And perhaps not surprisingly, each approach contains both advantages and disadvantages. Some triumphs of the different approaches include their success at predicting desired outcomes with relative accuracy, their ease of administration, and the information they yield for the user. Some of the limitations of these procedures include the problem of upscaling dynamic assessments to make them group administered and the problems associated with labeling individuals who are identified or misidentified as high potential. With the range of different approaches to measuring aptitude that currently exist and the likelihood that further techniques will be developed that align with different perspectives on the nature of aptitude, there is good reason to be optimistic about the future for aptitude assessment.

References

Beckmann, J. F., & Guthke, J. (1995). Complex problem solving, intelligence status, and learning ability. In J. Funke & P. Frensch (Eds.), *Complex problem solving: The European perspective* (pp. 177–200). Hillsdale, NJ: Erlbaum.

Beckmann, J. F., & Guthke, J. (1999). *Assessing reasoning ability*. Gottingen, Germany: Hogrefe.

Binet, A. (1916/1905). *New methods for the diagnosis of the intellectual level of subnormals. Classics in the history of psychology*. Retrieved from http://psychclassics.yorku.ca/Binet/binet1.htm

Birney, D. (2003). Mediating the impact of mediated learning: Review of "Experience of mediated learning: An impact of Feuerstein's theory in education and psychology" by A. Kozulin and Y. Rand (Eds.). *Contemporary Psychology: APA Review of Books, 48*, 677–679.

Birney, D., & Stemler, S. E. (2007). Intelligence quotient. In N. Salkind (Ed.), *Encyclopedia of measurement and statistics* (Vol. 2, pp. 473–476). Thousand Oaks, CA: Sage.

Brody, N. (1992). *Intelligence* (2nd ed.). San Diego, CA: Academic Press.

Brody, N. (2000). History of theories and measurements of intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 16–33). New York, NY: Cambridge University Press.

Brown, A. L., & French, L. A. (1979). The zone of potential development: Implications for intelligence testing in the year 2000. *Intelligence, 3*, 255–273. doi:10.1016/0160-2896(79)90021-7

Budoff, M. (1987). The validity of learning potential assessment. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 54–81). New York, NY: Guilford Press.

Camilleri, B. (2005). Dynamic assessment and intervention: Improving children's narrative abilities. *International Journal of Language and Communication Disorders, 40*, 240–242.

Campione, J. C., & Brown, A. L. (1987). Linking dynamic testing with school achievement. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 82–115). New York, NY: Guilford Press.

Carlson, J., & Wiedl, K. H. (1992). The dynamic assessment of intelligence. In H. C. Haywood & D. Tzuriel (Eds.), *Interactive assessment* (pp. 167–186). New York, NY: Springer-Verlag.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytical studies*. New York, NY: Cambridge University Press. doi:10.1017/CBO9780511571312

Cohen, R. J., & Swerdlik, M. E. (2005). *Psychological testing and assessment: An introduction to tests and measurements* (6th ed.). Boston, MA: McGraw-Hill.

Corno, L., Cronbach, L. J., Kupermintz, H., & Lohman, D. F. (2001). *Remaking the concept of aptitude: Extending the legacy of Richard E. Snow*. New York, NY: Routledge.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671–684. doi:10.1037/h0043943

Cronbach, L. J. (1967). How can instruction be adapted to individual differences? In R. M. Gagne (Ed.), *Learning and individual differences* (pp. 23–39). Columbus, OH: Merrill.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods*. New York, NY: Irvington.

Dewey, J. (1916). *Democracy and education*. New York, NY: Macmillan.

Dewey, J. (1938). *Experience and education*. New York, NY: Macmillan.

DuBois, P. H. (Ed.). (1947). *The classification program* (Army Air Forces Psychology Program Research Reports, No. 2). Washington, DC: U.S. Government Printing Office.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Random House.

Eden, D., & Shani, A. (1982). Pygmalion goes to boot camp: Expectancy, leadership, and trainee performance. *Journal of Applied Psychology, 67*, 194–199. doi:10.1037/0021-9010.67.2.194

Elliott, J. G. (2003). Dynamic assessment in educational settings: Realising potential. *Educational Review, 55*, 15–32. doi:10.1080/00131910303253

Erdley, C. A., & Dweck, C. S. (1993). Children's implicit personality theories as predictors of their social judgments. *Child Development, 64*, 863–878. doi:10.2307/1131223

Fabio, R. A. (2005). Dynamic assessment of intelligence is a better reply to adaptive behavior and cognitive plasticity. *Journal of General Psychology, 132*, 41–66. doi:10.3200/GENP.132.1.41-66

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, England: Freund.

Feuerstein, R., Klein, P. S., & Tannenbaum, A. J. (Eds.). (1991). *Mediated learning experience (MLE): Theoretical, psychosocial, and learning implications*. London, England: Freund.

Feuerstein, R., Rand, Y., & Hoffman, M. B. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device-theory, instruments, and techniques*. Baltimore, MD: University Park Press.

Gervey, B. M., Chiu, C., Hong, Y., & Dweck, C. S. (1999). Differential use of person information in decisions about guild versus innocence: The role of implicit theories. *Personality and Social Psychology Bulletin, 25*, 17–27. doi:10.1177/0146167299025001002

Gregory, R. J. (2007). *Psychological testing: History, principles, and applications* (5th ed.). Boston, MA: Pearson.

Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin? *Journal of Learning Disabilities, 42*, 111–132. doi:10.1177/0022219408326207

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin, 124*, 75–111. doi:10.1037/0033-2909.124.1.75

Guilford, J. P. (1967). *The nature of human intelligence*. New York, NY: McGraw-Hill.

Guthke, J. (1992). Learning tests: The concept, main research findings, problems, and trends. *Learning and Individual Differences, 4*, 137–151. doi:10.1016/1041-6080(92)90010-C

Guthke, J., Beckmann, J. F., & Dobat, H. (1997). Dynamic testing-problems, uses, trends, and evidence of validity. *Educational and Child Psychology, 14*, 17–32.

Hamers, J. H. M., Hessles, M. G. P., & Pennings, A. H. (1996). Learning potential in ethnic minority children. *European Journal of Psychological Assessment, 12*, 183–192. doi:10.1027/1015-5759.12.3.183

Haywood, H. C., & Lidz, C. (2007). *Dynamic assessment in practice*. New York, NY: Cambridge University Press.

Heslin, P. A. (2009). "Potential" in the eye of the beholder: The role of managers who spot rising stars. *Industrial and Organizational Psychology, 2*, 420–424. doi:10.1111/j.1754-9434.2009.01166.x

Heslin, P. A., Latham, G. P., & VanderWalle, D. (2005). The effect of implicit person theory on performance appraisals. *Journal of Applied Psychology, 90*, 842–856. doi:10.1037/0021-9010.90.5.842

Hessles, M. G. P., & Hamers, J. H. M. (1993). A learning potential test for ethnic minorities. In J. H. M. Hamers, K. Sijtsma, & A. J. J. M. Ruijssenaars (Eds.), *Learning potential assessment: Theoretical methodological, and practical issues* (pp. 285–311). Lisse, the Netherlands: Swets & Zeitlinger.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*, 253–270. doi:10.1037/h0023816

Jeltova, I., Birney, D., Fredine, N., Jarvin, L., Sternberg, R. J., & Grigorenko, E. L. (2007). Dynamic assessment as a process-oriented assessment in educational settings. *Advances in Speech Language Pathology, 9*, 273–285. doi:10.1080/14417040701460390

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin and Review, 9*, 637–671. doi:10.3758/BF03196323

Kaplan, R. M., & Saccuzzo, D. P. (2009). *Psychological testing: Principles, applications, and issues* (7th ed.). Belmont, CA: Wadsworth.

Kyllonen, P. C. (2005, September). The case for noncognitive assessments. *ETS R&D Connections*, 1–7.

Kyllonen, P. C., Roberts, R. D., & Stankov, L. (Eds.). (2008). *Extending intelligence: Enhancement and new constructs*. New York, NY: Erlbaum.

Lemann, N. (2000). *The big test: The secret history of the American meritocracy*. New York, NY: Farrar, Straus, & Giroux.

Lidz, C., & Elliott, J. G. (2000). *Dynamic assessment: Prevailing models and applications*. Greenwich, CT: Elsevier.

Matthew, C. T., Beckmann, J. F., & Sternberg, R. J. (2008). *Development of a test battery to assess mental flexibility based on Sternberg's theory of successful intelligence*. (Tech. Rep. No. 1222). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Matthew, C. T., & Stemler, S. E. (2008). *Exploring pattern recognition as a predictor of mental flexibility* (Final report submitted to the Army Research Institute for the Behavioral and Social Sciences for Project Award #W91WAW-07-C-00). Middletown, CT: Wesleyan University.

Mone, E. M., Acritani, K., & Eisinger, C. (2009). Take it to the roundtable. *Industrial and Organizational Psychology, 2*, 425–429. doi:10.1111/j.1754-9434.2009.01167.x

Mueller-Hanson, R. A., Swartout, E., Hilton, R., & Nelson, J. (2009, June). *Proof of concept research for developing adaptive performance: Validation plan*. Arlington, VA: Personnel Decisions Research Institutes.

Olswang, L. B., & Bain, B. A. (1996). Assessment information for predicting upcoming change in language production. *Journal of Speech and Hearing Research, 39*, 414–423.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9*, 105–119.

Plaks, J. E., Stroessner, S. J., Dweck, C. S., & Sherman, J. W. (2001). Person theories and attention allocation: Preferences for stereotypic versus counterstereotypic information. *Journal of Personality and Social Psychology, 80*, 876–893. doi:10.1037/0022-3514.80.6.876

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*, 299–324. doi:10.1207/S15327043HUP1504_01

Rosenthal, R., & Jacobson, L. (1992). *Pygmalion in the classroom* (Expanded ed.). New York, NY: Irvington. (Original work published 1968)

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274. doi:10.1037/0033-2909.124.2.262

Schmitt, N., Oswald, F. L., & Gillespie, M. A. (2005). Implications of broadening the performance domain for the prediction of academic success. In W. F. Camara & E. Kimmel (Eds.), *Choosing students: Higher education admission tools for the 21st century* (pp. 195–214). Mahwah, NJ: Erlbaum.

Schmitt, N., Oswald, F. L., Kim, B. H., Imus, A., Drzakowski, S., Friede, A., & Shivpuri, S. (2007). The use of background and ability profiles to predict college student outcomes. *Journal of Applied Psychology, 92*, 165–179. doi:10.1037/0021-9010.92.1.165

Silzer, R., & Church, A. H. (2009a). The pearls and perils of identifying potential. *Industrial and Organizational Psychology, 2*, 377–412. doi:10.1111/j.1754-9434.2009.01163.x

Silzer, R., & Church, A. H. (2009b). The potential for potential. *Industrial and Organizational Psychology, 2*, 446–452. doi:10.1111/j.1754-9434.2009.01172.x

Snow, R. E. (1977). Individual differences and instructional theory. *Educational Researcher, 6*(10), 11–15.

Snow, R. E. (1978). Theory and method for research on aptitude processes. *Intelligence, 2*, 225–278. doi:10.1016/0160-2896(78)90019-3

Snow, R. E. (1992). Aptitude theory: Yesterday, today, and tomorrow. *Educational Psychologist*, *27*, 5–32. doi:10.1207/s15326985ep2701_3

Spear, L. C., & Sternberg, R. J. (1987). Teaching styles: Staff development for teaching thinking. *Journal of Staff Development*, *8*(3), 35–39.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, *15*, 201–293. doi:10.2307/1412107

Stemler, S. E. (2009). *The measurement of adaptability in military personnel.* Orlando: Institute of Simulation and Training, University of Central Florida.

Stemler, S. E. (2012). What should university admissions tests predict? *Educational Psychologist*, *47*, 5–17. doi: 10.1080/00461520.2011.611444

Stemler, S. E., Grigorenko, E. L., Jarvin, L., & Sternberg, R. J. (2006). Using the theory of successful intelligence as a basis for augmenting AP exams in psychology and statistics. *Contemporary Educational Psychology*, *31*, 75–108.

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. doi:10.1016/j.cedpsych.2009.04.001

Sternberg, R. J. (1985). *Beyond IQ.* New York, NY: Cambridge University Press.

Sternberg, R. J. (1995). *In search of the human mind.* Orlando, FL: Harcourt Brace.

Sternberg, R. J. (1997). *Successful intelligence.* New York, NY: Plume.

Sternberg, R. J. (2003a). *Sternberg triarchic abilities test.* Unpublished test.

Sternberg, R. J. (2003b). *Wisdom, intelligence, and creativity synthesized.* New York, NY: Cambridge University Press. doi:10.1017/CBO9780511509612

Sternberg, R. J. (2005). The theory of successful intelligence. *Interamerican Journal of Psychology*, *39*, 189–202.

Sternberg, R. J. (2009, Fall). WICS: A new model for liberal education. *Liberal Education*, *95*(4), 20–25.

Sternberg, R. J., Bonney, C. R., Gabora, L., Karelitz, T., & Coffin, L. (2010). Broadening the spectrum of undergraduate admissions. *College and University*, *86*, 2–17.

Sternberg, R. J., Ferrari, M., Clinkenbeard, P., & Grigorenko, E. L. (1996). Identification, instruction, and assessment of gifted children: A construct validation of a triarchic model. *Gifted Child Quarterly*, *40*, 129–137. doi:10.1177/001698629604000303

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J., Snook, S., Williams, W. M., . . . Grigorenko, E. L. (2000). *Practical intelligence in everyday life.* New York, NY: Cambridge University Press.

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing.* New York, NY: Cambridge University Press.

Sternberg, R. J., Grigorenko, E. L., Ferrari, M., & Clinkenbeard, P. (1999). A triarchic analysis of an aptitude–treatment interaction. *European Journal of Psychological Assessment*, *15*, 3–13. doi:10.1027//1015-5759.15.1.3

Sternberg, R. J., Grigorenko, E. L., & Zhang, L.-F. (2008a). A reply to two stylish critiques. *Perspectives on Psychological Science*, *3*, 516–517. doi:10.1111/j.1745-6924.2008.00092.x

Sternberg, R. J., Grigorenko, E. L., & Zhang, L.-F. (2008b). Styles of learning and thinking matter in instruction and assessment. *Perspectives on Psychological Science*, *3*, 486–506. doi:10.1111/j.1745-6924.2008.00095.x

Sternberg, R. J., & Lubart, T. I. (1995). *Defying the crowd: Cultivating creativity in a culture of conformity.* New York, NY: Free Press.

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. doi:10.1016/j.intell.2006.01.002

Sternberg, R. J., & Williams, W. M. (1996). *How to develop student creativity.* Alexandria, VA: Association for Supervision and Curriculum Development.

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago, IL: University of Chicago Press.

Today's Military. (2012). *ASVAB test.* Retrieved from http://www.todaysmilitary.com/before-serving-in-the-military/asvab-test

U.S. Department of Education. (2010). *Statement on National Governors Association and State Education Chiefs Common Core Standards.* Retrieved from http://www.ed.gov/news/press-releases/statement-national-governors-association-and-state-education-chiefs-common-core

Vernon, P. E. (1950). *The structure of human abilities.* London, England: Methuen.

Vygotsky, L. W. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press. (Original work published 1934)

Weinstein, R. S. (2002). *Reaching higher: The power of expectations in schooling.* Cambridge, MA: Harvard University Press.

Yost, P. R., & Chang, G. (2009). Everyone is equal, but some are more equal than others. *Industrial and Organizational Psychology*, *2*, 442–445. doi:10.1111/j.1754-9434.2009.01171.x