CrossMark

Context-Sensitive Cognitive and Educational Testing

Robert J. Sternberg[1,2]

Abstract This article reviews four interrelated approaches to reducing an inequitable gap in cognitive and educational test scores between individuals of a dominant culture and individuals of other cultures or subcultures. These approaches include (a) use of broader measures, (b) performance- and project-based assessments, (c) direct measurement of knowledge and skills relevant to environmental adaptation, and (d) dynamic assessment. It is concluded that when appropriate assessment is done that recognizes students' diverse cultural and social backgrounds, equity can increase, predictive validity of cognitive and educational tests can increase, and at the same time, racial/ethnic/culture differences can decrease.

Intelligence is often defined as the ability to adapt to the environment (see Sternberg and Kaufman 2011). Suppose you enter a room to take a test that you know will measure your adaptive abilities. It already has been explained to you that the test will be performance-based. Perhaps you think of the performance tests on the Wechsler Adult Intelligence Scales (WAIS) or Wechsler Intelligence Scales for Children (WISC) so you are not particularly worried. You have been through this before.

Once you enter the room, you are handed a Remington 700 SPS Tactical AAC-SD, Bolt Action, .308 Winchester, 20″ inch barrel, 4 + 1 rounds. You are told that, once you go outdoors into the nearby terrain, you will have 3 h to shoot as much game as you can. The examiner cracks a smile and says you are lucky, because not many years ago, the only equipment you would have been given was a bow-and-arrow. He wishes you good luck, opens the back door to the vast exterior, and says "Go." You are on your own.

For people who live "off the grid," so to speak, this (imagined) hunting test might be an appropriate assessment of intelligence as adaptation to the environment, as might be a test of

✉ Robert J. Sternberg
robert.sternberg@cornell.edu

[1] Cornell University, B44 MVR, Ithaca, NY, USA

[2] Department of Human Development, College of Human Ecology, Ithaca, NY 14853, USA

gathering skills or of skills in using natural herbal medications to combat illnesses (see Laboratory of Comparative Human Cognition 1982; Serpell 2000; Sternberg 2004, 2007). There are people who live off the grid in the USA, although not so many. But worldwide, there are many, sometimes in remote or extremely rural areas, and sometimes in war or deprivation zones where foraging for food is an everyday challenge, not just for oneself, but also for family members. For individuals growing up in some of these off-the-grid environments—present and past—sitting down in a classroom and taking a paper-and-pencil intelligence test might seem as strange and unnatural as the hunting test possibly might to you.

In speaking of "cognitive and educational testing" this article refers to tests traditionally referred to as "intelligence tests," but also to tests that are in many samples moderately to highly correlated with such tests, such as the SAT (Frey and Detterman 2004), ACT (Koenig et al. 2008), and various educational achievement tests that emphasize material taught directly in Western schooling. Indeed, Sternberg has argued that tests such as IQ tests, SATs, and ACTs, and achievement tests can be viewed on a continuum in terms of their requirements for knowledge taught in Western schooling, as almost all of the tests require such knowledge in varying degrees (Sternberg 1998).

This essay will review the work of various investigators, but will focus on my colleagues' and my work as an example of a research program that addresses broad issues of context-sensitive cognitive testing. A full review of relevant literature would go way beyond the scope of a journal review article. Many other sources are cited in this article that provide additional references as well.

One could ask whether it is possible "that ability [can] be assessed independently of the particular national or eco-cultural context in which…success will be manifested." This was the noble goal of tests of abstract reasoning, such as the Raven Progressive Matrices (Raven et al. 2003) or what was called the "Cattell Culture Fair Intelligence Tests" (Cattell et al. 1973). For many years, most if not almost all psychologists believed such tests to be truly culture-fair, or as Cattell originally thought, culture-free (Cattell 1949). But the Flynn effect (Flynn 1987) shows the greatest IQ gains in tests once thought to be culture-fair or even culture-free—those of fluid abilities—suggesting that even scores on tests that seem on the surface to be free of influences of Western schooling in fact are highly dependent on such schooling. *There are no culture-fair or culture-free tests* (Sternberg 1990). Psychologists only thought there were such tests because they did not fully recognize their own cultural presuppositions, much as people typically do not recognize the accent in their own voice when they speak, although they can recognize other people's accents. Their own voice sounds unaccented, just as one's own culture can just seem like "the way things are."

Some researchers (Serpell 2017) have spent at least a portion of their careers trying to design cognitive and educational tests that are appropriate for the environments of individuals who grow up in other than traditional Western environments (see Sternberg and Grigorenko 2004a) and for whom Western tests measure skills other than those for which their environments have socialized them (Sternberg and Suben 1986). One of the first to recognize the at least partial cultural relativity of many cognitive skills was John Berry (see, e.g., Berry 1974; Berry and Irvine 1986; Berry et al. 1992). Moreover, Serpell (2017) has shown how cognitive testing in rural Africa can reveal far greater levels of abilities in children if the testing takes into account their indigenous skills. Saxe (1991, 2012) has shown that children in Papua New Guinea and elsewhere have sophisticated mathematical understandings that would not be tapped by conventional tests. Nuñes et al. (1993) have demonstrated that Brazilian street children have complex knowledge of how to do mathematics on the street, but that knowledge

transfers poorly or not at all to performance on standardized tests (see also Ceci and Roazzi 1994). Jean Lave (1988) has showed how housewives who could do complex computations in the context of supermarket shopping could not do analogous problems in the context of a school classroom. Barbara Rogoff (1991, 2003) and Patricia Greenfield and her colleagues (Greenfield et al. 1997) have shown how Mayan children learn complex skills relevant to their environments that would leave middle-class Western children far beyond. Mary Gauvain (Gauvain 2013; Gauvain and Munroe 2012, 2013) further has demonstrated how sociocultural contexts can have an effect on a wide variety of educational performances.

Rogoff et al. (2017), Cole (2017), and others such as Lave (1988) and Greenfield et al. (1997) have emphasized the importance of cultural practice for understanding cognition in its cultural context. In particular, they take a strengths-based approach and emphasize how individuals are socialized to work together. Rogoff et al., for example, have studied how sophisticated collaborations are embedded in the way indigenous peoples in the Americas solve problems, and in general, perform tasks. These skills are purposely socialized so that individuals in a community learn how to collaborate with each other; they also learn from other communities to expand the ways in which they get things done. In other words, through communities of practice, they can accomplish tasks that urbanized Westerners would never learn to do (such as complex forms of weaving), and if they did learn, probably could never achieve a high level of skill because of their individualistic way of approaching the tasks.

The contribution this article is to go beyond these studies and other excellent work by placing some of what psychology and education understand about context sensitivity into a single theoretical framework, that of the theory of successful intelligence (Sternberg 1985c, 1993, 2004; Sternberg and Grigorenko 2004b; Sternberg and Hedlund 2002). My colleagues and I further attempted to show how this theory can help understand cognitive and educational performances in a wide variety of cultural and subcultural contexts.

The theory is based on the notion that intelligence is not merely what intelligence tests measure. Rather, what is called "successful intelligence" is the ability to choose and sometimes re-choose a life course that is prosocial, personally meaningful, and self-fulfilling, and that enables one to capitalize on one's strengths and to compensate for or correct one's weaknesses, in order to adapt to, shape, and select environments. According to this theory, therefore, typical societal measures of success are useful in the same way that average heights or average incomes are. Such averages may reflect many people, some people, or no people at all. (If, for example, among four people, two are extremely poor and two are extremely rich, their mean income may be an "average," but actually reflects none of their individual incomes.) Societally defined measures of success may be important to some people, but they are unlikely to be important to all people, nor should they be. Otherwise, in stagnant societies, no one would exhibit the creativity to move the societies away from their own stagnancy.

The theory of successful intelligence suggests that schools not only should teach students what societally valued outcomes are but also should aid students in developing their own set of valued outcomes, which may overlap with the societal ones but also may contain many unique elements. The schools should teach not only what outcomes the students' own societies value but also the outcomes other societies value. Only in this way can students learn that measures of success are not universal—that they need to be adapted to the society, to groups within the society, and to the specific individuals within those groups.

The underlying components for successful intelligence are creative, analytical, practical, and wisdom-based skills, as discussed further below. At the base of the theory of successful intelligence is a set of information-processing components that can be used in solving

problems of different types, including creative, analytical, practical, and wisdom-based problems (Sternberg 1985a).

There are three kinds of components—metacomponents, performance components, and knowledge-acquisition components. Consider each in turn.

Metacomponents, or executive processes, are used to (a) recognize the existence of a problem, (b) define the problem, (c) mentally represent information about the problem, (d) formulate a strategy to solve the problem, (e) monitor problem solving while it is ongoing, and (f) evaluate problem solving after it is done. An example of metacomponents at work would be writing a psychology paper, where one has to figure out that there is a problem to address, define exactly what the problem is, organize (mentally represent) information about that problem, and so forth.

Performance components execute the instructions of the metacomponents. There are many of these, but among the more important ones are (a) inferring the relations between concepts, (b) mapping higher order relations between simple relations, (c) applying relations that have been inferred, and (d) justifying one solution among those given as the best, if not ideal solution (Sternberg 1983). An example of the use of performance components would be the analogy Washington: 1: Lincoln: (13, 14, 15, 16), where inference would be used to figure out the relation between "Washington" and "1"; mapping would be used to see the higher order relation between "Washington" and "1," on the one hand, and between "Lincoln" and an answer option, on the other; application would be used to apply from "Lincoln" the relation that was inferred between "Washington" and "1"; and justification would be used to choose the best option (16), even if it is not an ideal option. In this case, the problem is about the ordinal number of Lincoln's presidency, not, for example, about the U.S. coin or currency value on which Lincoln appears.

Knowledge-acquisition components are used to figure out how to solve problems, for example, (a) selective encoding, or deciding what information in a problem is relevant; (b) selective comparison, or deciding what information stored in long-term memory is relevant to problem solution; and (c) selective combination, or deciding how to combine elements that have been selectively encoded and compared to reach a solution. An example would be "Black and blue socks in a drawer are mixed in a ratio of 4:5. How many socks does one have to take out of the drawer to be assured of having a pair of the same color? In this case, one uses selective encoding to recognize that what is relevant to solution is the fact that there are just two colors of socks in the drawer (and not the fact that they are mixed in a ratio of 4:5 or that they happen to be black and blue). One uses selective comparison to figure out what happens when one lays socks of two different colors on a bed or drawer or wherever. And one uses selective combination to figure out that, regardless of ratio or color, the first sock must be black or blue, the second must be black or blue—and if the same color as the first one, there is a match—and the third must provide a match because it will match either the blue or black sock(s) already extracted from the drawer.

In a previous article (Sternberg 2004), I describe four models by which this theory might be applied. Consider each in turn.

In model I, the nature of intelligence is the same across cultures, as are the tests used to measure intelligence. Model I is the theoretical position of psychologists such as Jensen (1998). This position holds that the nature of intelligence is the same across various cultures and that this nature can be assessed identically (using appropriate translations of text, where necessary) without regard to culture.

Model II represents a difference in the nature of intelligence but no difference in how intelligence manifests itself or in the instruments used to measure it. The measures used to assess intelligence are the same across cultures, but the outcomes obtained from using those measures are structurally different as a function of the culture being investigated. This approach is close to that taken by Nisbett (2004). Nisbett found that the same tests given in different cultures suggested that, across cultures, people think about problems in different ways. Thus, Nisbett uses essentially the same tests to elicit different ways of thinking across cultural groups Although intellectual functioning differs across cultural contexts, the same measures can be used to assess that functioning.

In model III, the dimensions or aspects of intelligence are the same, but the manifestations of intelligence and the instruments used to measure intelligence are not. In this model, measurement processes for a given attribute must be emic, that is, derived from the context of the culture being studied rather than from outside it. This is not to say that the same instruments cannot be used across cultures; but when they are, the psychological meanings to be assigned to the scores will differ from one culture to another. This is the position taken in this article and in my earlier work (e.g., Sternberg 1990).

In model IV, both the instruments and the ensuing dimensions or aspects of intelligence are different as a function of the culture under investigation. This position embraces the radical cultural-relativist position (Berry 1974) that intelligence can be understood and measured only as an indigenous construct within a given cultural context. It also embraces the position of Sarason and Doris (1979), who view intelligence largely as a cultural invention. In other words, nothing about intelligence is necessarily common across cultures.

Model III, the model adopted in this article, implies that the actions that constitute intelligence can vary widely over cultures. There are few common threads between, say, a child growing up in Scarsdale, NY, and one growing up in rural villages near Kisumu, Kenya, or even the black ghettoes in Harlem, not so many miles from Scarsdale. Tasks measuring intelligence would need to be rather different, simply because the adaptive requirements of the various cultures or subcultures are so different. But the processes of intelligence are the same: in all cases, individuals need the metacomponents, performance components, and knowledge acquisition components underlying the intelligent behavior, which are the same. Thus, when I describe cognitive tasks in this article, my argument is not that each task is relevant in each culture or subculture. Rather, the components underlying creative, analytical, practical, and wisdom-based skills are the same, but how they are manifested and how they should be measured are likely to differ.

The work described here in this article seeks to be relevant not just to individuals who grow up in off-the-grid environments but also to children and adults who grow up in inner city environments, rural environments, or other culturally diverse environments where traditional Western cognitive testing either is not familiar or not important—that is, motivating—to them. Moreover, expanded testing of the kind described in this article can be useful as well to individuals who grow up in conventional Western middle-class environments because it may elicit skills beyond the knowledge-based and abstract analytical skills that conventional standardized tests measure. For some middle-class children, their strengths are in knowledge and skills other than those conventionally tested (see also Gardner 2011).

Individuals who are members of groups that historically have not performed well on cognitive and educational tests sometimes do poorly not only because of reduced relevance of the tests to their background, but also because of stereotype threat (e.g., Steele 1997; Steele and Aronson 1995). Such individuals may expect members of their group to perform poorly on

standardized tests, such that the expectation becomes a self-fulfilling prophecy. Following Steele and his colleagues, one might reduce stereotype threat by minimizing the importance of stereotype threat by trying to make it non-salient in the testing situation (e.g., "This is not a high-stakes test of your abilities") or by ensuring, to the extent possible, that membership in stereotyped groups is not primed in the testing situation.

The larger problem, however, is not with the testing situation, but rather with the stereo-types societies perpetrate and even encourage. For example, one of the strongest stereotype threats is age (Hess et al. 2003). Unlike some stereotype threats, this is one everyone will face in his or her life, at some time or another. Societies are not likely to remove these stereotype threats, nor are they always aware of them. The best solution may be to teach students about them directly and to show how they can damage performance. This procedure, unfortunately, will not "solve" the problem, and risks make the stereotype threat more salient than it was before. The advantage of teaching about stereotype threat, however, is that at least people will be aware of the situation, and perhaps those evaluating test results will also take into account potential effects of stereotype threats of various kinds.

This article describes four interrelated approaches that can be used in order better to meet the needs of individuals who may have important adaptive cognitive skills that are bypassed. The approaches are (a) broader cognitive measures, (b) performance- and project-based assessments, (c) direct measurement of knowledge and skills relevant to environmental adaptation, and (d) dynamic testing. These four interrelated techniques all can help elucidate knowledge and skills that might not be uncovered in standard cognitive testing, and that are relevant for intelligence as adaptation to the environment.

Broader Conceptualizations of Cognitive Skills

Typical measures of cognitive and educational skills, ranging from conventional achievement tests in high school to tests such as the SAT and ACT, assess primarily memory, knowledge, and analytical-reasoning skills (Sternberg 1997a, b, 2003, 2010). These tests do not attempt to assess broader cognitive skills, perhaps because such skills are harder to measure or perhaps because of historical precedent. Perhaps, then, it is time to consider new and broader approaches to assessment.

There have been many attempts to measure broader cognitive skills than those assessed by conventional tests of academic skills. Different bases have been used to try to broaden assessments. Indirectly, at least, many of these approaches hark back to the work of Lesser et al. (1965). These researchers discovered that members of different socially defined groups showed different patterns of scores on mental-ability tests. The study was ground-breaking because it suggested at least one reason that different groups may show different average scores on tests of intelligence. The groups may have different patterns of developed cognitive skills, with the content of tests balanced toward cognitive skills that favor some groups over others.

Another body of work that is relevant is that of John Ogbu (2003, 2008), who found that individuals from minority groups whose ancestors were forcibly repatriated from their home countries (e.g., African-Americans) tend to adapt less well in many ways than do majority-group members or minority-group members who chose their destination. The descendants of those who were repatriated bear psychological scars that others do not have to bear, and their adaptation suffers accordingly.

Cronbach (1957), in his presidential address to the American Psychological Association, made a plea to unite the two disciplines of scientific psychology (experimental and psychometric). Cronbach pointed out that both experimental and differential (psychometric) approaches in psychology have a lot to contribute to our understanding of human beings, but oddly, at least in 1957, the approaches rarely had been integrated or used in some kind of synergistic way.

Cronbach's plea has been addressed in a number of ways (see, e.g., Gardner 2011; Hunt 2011; Mackintosh 2011; Sternberg 1985b, 1988a, b; Sternberg and Kaufman 2011), including the ways described below. In the work to be cited, researchers use a variety of approaches, including the differential and experimental, but also the explicitly cultural, which seeks to understand differences in adaptive requirements across cultures. For example, in contemporary US culture, technological skills and skills in data retrieval using technological and other means are very important for success in schooling and thereafter. But in other cultures, hunting skills, fishing skills, gathering skills, artisanship, and other kinds of skills may be far more important, as they were in earlier times and still are today.

Although the work described here has been motivated by the theory of successful intelligence, other motivating theories have been used as well. For example, some broader assessments derive from the theory of multiple intelligences (Gardner 2011). This theory posits linguistic, logical-mathematical, naturalist, musical, spatial, interpersonal, intrapersonal, and bodily kinesthetic intelligences. Each intelligence is alleged to be independent. As an example, Krechevsky (1998) developed preschool assessments based on the theory that are performance-based and that engage the various intelligences. However, some of the assessments that have been formulated have been somewhat ill-defined and, moreover, it is debatable how well the theory has stood up empirically (Visser et al. 2006; but see also Schaler 2006). Whether the theory is correct or not in its present form, the skills it posits are useful in life and extend beyond the content of traditional tests of intelligence and cognitive skills. The theory needs further empirical validation, which will depend on the formation of construct-valid assessments that can be used for such purposes.

As mentioned, many of our tests derive from the theory of successful intelligence (Sternberg 1984, 1985a, 1997a, b, 2003, 2010). The basic idea of this theory is that intelligence comprises a related set of constructs—creative skills to generate new and useful ideas, analytical skills to evaluate the quality of these (and other) ideas, practical skills to implement the ideas and to persuade others of their usefulness, and wisdom-based skills to turn the ideas toward the attainment of a common good. Some of the earlier tests of the theory were based on a largely multiple-choice triarchic abilities test (Sternberg 1993), but more recent assessments have used performance- and project-based tests (e.g., Sternberg 2010). Some of these assessments are described later in this article.

Other assessments have branched out into personality and related constructs (Arteche et al. 2009; DeYoung 2011) or into emotional intelligence (Mayer et al. 2011). Goleman (1995), for example, has suggested that emotional intelligence has five components: self-awareness, self-regulation, motivation, empathy, and social skills. This definition is perhaps overly broad: certainly people, such as creative scientists or engineers, can be highly motivated but not high in emotional intelligence. Mayer and Salovey (1993) have suggested a model that seems better to capture the construct: accurately perceiving emotions in oneself and others, using emotions to facilitate thinking, understanding emotional meanings, and managing emotions. They have devised a test of emotional intelligence, the MSCEIT (Mayer Salovey Caruso Emotional Intelligence Test), which appears to have good predictive value to a wide range of behavior (Mayer et al. 2004; Mayer et al. 2003).

Why might groups from certain cultural or other environments fare better on broader tests than they do on narrower ones? There are at least two reasons, deriving from (a) explicit theories of intelligence and (b) implicit (folk) theories of intelligence.

Explicit theories of intelligence deal, or at least are supposed to deal with the challenges people face in adapting to their environments. Youth who grow up in challenging environments must, by necessity, develop creative and practical skills (Sternberg 1988b). For example, Abrams and Terry (2014) found that incarcerated young adult men, when released back into their former environments, need uniquely adapted creative and practical skills in order to survive back on the streets. The experience of these men is not typical of individuals in various challenging environments, but rather shows how one group—and a fairly large group, as many people in the USA are incarcerated and then released—tries to adapt to the changing circumstances of their environments. As mentioned above, creative and practical skills are relevant almost everywhere, but how they are manifested differs. In particular, rather than experiencing their late teens and early twenties as times of exploration and gradual attainment of independence, these individuals need to focus on economic and physical survival—especially avoiding dangers and diversion, taking calculated risks to maximize their chances of survival, and running and hiding. They have to hide not only from members of gangs other than their own, they need often to hide from the police and even, sometimes, from members of their own gang. For these men, to the extent intelligence is about adaptation—and literally the skills needed to survive—the skills tested on intelligence tests take a back seat to recognizing clear and present dangers in their environment. They simply do not have the time or the resources to focus on the skills that would lead to the development of high scores on tests of IQ, or SATs, or ACTs, or whatever. Their adaptive skills need to be directed toward tasks different from those that would tend to occupy middle-class youth on the threshold of adulthood. Measures of adaptive skills thus form one basis for testing intelligence broadly, particularly because intelligence is supposed to measure adaptive skills. Why not measure them directly rather than indirectly, as in a conventional intelligence test?

What is the relevance of the Abrams and Terry (2014) study to people who are not "under great social pressure to rejoin criminal groups of which they had been members at the time of their arrest"? To the millions of incarcerated individuals who are released from prison, the pressure to stay out of prison but also to rejoin former social groups is extremely life-relevant. To a reviewer of this article, the life circumstances understandably may seem quaint or far-fetched. But to these formerly incarcerated men, the circumstances of taking a standardized test with analogies or math problem about water lilies similarly may seem quaint and far-fetched.

These newly released individuals have to worry about not being reincarcerated and about how to deal with gang members who retain an abiding interest in them. For the most part, those who are released from prison hardly could care less about academic analogies or abstract mathematical problems on standardized tests that are completely divorced from the contexts of their lives. Their criteria for life success may be very different from those of individuals seeking top college or graduate- or professional-school placements. People adapt to their real-life circumstances, which may or may not correspond to those relevant to high performance on conventional standardized tests. That is why tests of practical intelligence are important—they show people's abilities to adapt to the real circumstances of their lives, not necessarily the circumstances that test-constructors happen to think are important (Gardner et al. 1994; Sternberg et al. 2000).

Implicit theories of intelligence deal with people's conceptions of intelligence, and these theories may differ from one culture to another, leading parents to emphasize different

intellectual skills in raising their children. For example, Chinese people in Taiwan include practical interpersonal and intrapersonal (self-understanding) skills as part of their conception of intelligence (Yang and Sternberg 1997). In other words, Taiwanese parents raise children to see their relations with others and with themselves as important to being intelligent. Rural Kenyan conceptions of intelligence encompass ethical/moral as well as cognitive skills (Grigorenko et al. 2001). Thus, what might count as a comprehensive assessment of intelligence could differ from one culture to another (Sternberg 2004). In one study, teachers' notions of what it means to be smart corresponded better to parental conceptions in Asian-American and European-American students than to parental conceptions of Latino-American parents. The teachers thus judged the Asian-American and European-American students as more intelligent, on average (Okagaki and Sternberg 1993). Thus, what constitutes intelligence, at least as perceived by different cultural or subcultural groups, may differ. Performance- and project-based assessments may be a way of measuring the varied skills believed to constitute intelligence across cultures.

Performance- and Project-Based Assessment

Most standardized tests rely heavily on multiple-choice format, at least in the USA (as well as many other countries). Some organizations, such as the Association of American Colleges and Universities (AAC&U), among other organizations, have recognized the importance of broader assessments, such as portfolio assessments (https://www.aacu.org/peerreview/2014/winter), but these efforts remain minority ones.

There are some data suggesting that the multiple-choice format may be non-ideal for assessing the abilities of students from a broad range of environments.

The Rainbow Project

One avenue for identifying gifted students is through college and university admissions. When universities make admissions decisions, the principal quantitative data they use typically are high school grade-point average and performance on standardized tests (Sternberg 2010). Is it feasible to devise psychometrically sound assessments that furnish increased prediction of college GPA (and others measures of success) beyond that obtained by existing measures, without destroying the cultural, ethnic, and other forms of diversity that render a university environment the kind of place in which students can interact with and learn from other individuals who differ from themselves in key respects? Put another way, can one devise assessments that assess people's differing gifts that are potentially apposite to success in the university and in life? And can one do so in a manner that does not merely echo students' racial or socioeconomic status (see Hunt and Carlson 2007; Sternberg et al. 2005)?

The Rainbow Project (Sternberg and the Rainbow Project Collaborators 2006; Sternberg et al. 2012) was created to improve college-admissions procedures. The Rainbow assessments were devised to supplement the SAT or ACT, but they also could supplement any conventional other conventional standardized test of cognitive skills or achievement. The theory of successful intelligence views cognitive skills and achievement as existing on a continuum. On this view, mentioned earlier, cognitive skills are in large part achieved rather than merely being innate (Sternberg 1998). This view is consistent with some other views as well (e.g., Cronbach 1990; Gardner 2011). The skills covered are the creative, analytical, practical, and wisdom-

based skills in the theory of successful intelligence. (Wisdom was not covered in the earlier version of the theory—Sternberg 1997a, b. It was added later—Sternberg 2003).

In the Rainbow Project, my colleagues and I (Sternberg and the Rainbow Project Collaborators 2006) collected data from 1013 students at 15 US institutions, including 8 four-year colleges, 5 two-year colleges, and 2 high schools. The analytical items were of the kinds one would find on a traditional cognitive-abilities test. The multiple-choice tests included figuring out meanings of new words from context, number series, and figural reasoning.

Creative thinking skills were assessed via both multiple-choice items and performance-based items. An example of a creative multiple-choice item was verbal analogies preceded by counterfactual premises (e.g., suppose that villains were lovable). Test-takers then had to solve the analogies as though the counterfactual premises of the analogies were true (see Sternberg and Gastel 1989a, b). Creative thinking skills also were assessed via open-ended measures, such as writing short stories with unusual titles, like "The Professor Disappeared," captioning cartoons, and creative oral story telling in response to a visual collage, for example, of musicians or athletes. The project also used various types of multiple-choice items to assess practical skills, for example, solving everyday problems presented verbally, everyday mathematics (such as buying tickets to a baseball game), and using maps to plan routes.

Examples of creative performance-based tasks in the Rainbow Project were to write very short stories with suggested titles, such as "3516" or "It's Moving Backward" (see Lubart and Sternberg 1995). One type of performance-based practical item presented short videos for which test-takers saw scenarios that were incomplete: The movies involved common problems faced by college students. One item, for example, involved a student asking another student for help when the other student was on a "hot" date with a girlfriend. Test-takers then were asked to judge the quality of response options with respect to each situation.

There were no strict time limits for completing the tests; however, the test proctors were told to allow roughly 70 min per testing session. Creativity in the Rainbow (and the subsequent Kaleidoscope) Project was measured by considering both the novelty (or originality) and the quality of responses. Practicality was assessed based on the feasibility of the products considering both human and material resources. Reliabilities of measures were generally high. For example, reliability was .94 for written stories and .97 for oral stories (see Stemler et al. 2006).

The first research question was whether the assessments of the Rainbow Project actually measured separable analytical, creative, and practical skills, rather than simply the general (g) factor characterizing most conventional tests of cognitive skills. Factor analysis, which decomposes correlations between all possible pairs of tests, was used to answer this question. Three meaningful factors emerged: practical skills as measured by the practical performance tests, creative skills as measured by the creative performance tests, and analytical skills as measured by *all* of the multiple-choice tests (including not just the analytical ones, but also the creative and practical ones). In particular, when our Rainbow assessments were factor analyzed, the loadings on factor I (creative) were .57 and .79 for written and oral stories. All the other loadings were totally trivial. The loadings on factor III (practical) were .52 for movies, 1.00 for college-life common sense, and .92 for everyday common sense; everything else was trivial. For factor II (analytical/multiple-choice/general intelligence), the loadings were .73 for the multiple-choice creative, .80 for the multiple-choice analytical, and .81 for the multiple-choice practical items; all other loadings were trivial. Put another way, the multiple-choice tests, regardless of what they were supposed to measure, produced an analytical or "general" factor. Thus, method of assessment proved to be critical.

Why did multiple-choice tests cluster into a single general factor? One cannot be certain, but it appears that they all required (a) convergent thinking with (b) well-structured problems, (c) that were presented to the students rather than discovered by them, and that were (d) largely academic in nature. Multiple-choice tests are probably not the answer (Sternberg 2015a, b; Sternberg and the Rainbow Project Collaborators 2006) to diversifying assessment.

College admissions officers are not interested in whether new measures simply predict college academic success. Rather, they are interested in incremental validity—the extent to which new measures predict school success beyond those measures that are currently being used, such as the SAT and high school grade-point-average (GPA). To assess the incremental validity of the Rainbow measures above and beyond the SAT/ACT in predicting GPA, we conducted hierarchical regressions that added analytical, creative, and practical assessments to SAT and high school GPA.

With regard to simple correlations, the SAT-V, SAT-M, high school GPA, and the Rainbow measures all predict first-year year GPA. But how did the Rainbow measures fare with respect to incremental validity? The SAT-V, SAT-M, and high school GPA were placed into the first step of the prediction equation because these are the standard measures used today to predict college academic performance. Only high school GPA contributed uniquely to prediction of undergraduate GPA. However, placing the Rainbow measures into a next step of the hierarchical multiple regression essentially doubled prediction (percentage of variance accounted for in the criterion) versus the SAT alone.

To be more precise, the prediction (squared correlations) of first-year college GPA by SAT was .098 and with the addition of our own multiple-choice items was .099. In other words, we found, as did Spearman (1927), that adding more analytical measures to prediction, no matter how different they look, makes no difference to prediction—Spearman referred to this as the "indifference of the indicator." When we added the practical performance items to SAT, the prediction went up to .129. When we added the creative performance items to SAT, the prediction went up to .186, and when we added analytical, practical, and creative, the prediction was .209. If we added high school GPA as well, the prediction was at .248. (The squared correlations are lower than might be expected because they represent grading systems at a wide variety of colleges, ranging from unselective to highly selective. We did not correct GPAs for college selectivity in the original data report. When we did such a correction, the correlations were quite a bit higher.)

Thus, the Rainbow assessments substantially increased the level of prediction beyond that resulting from SATs on their own. The results also indicate the power of high school GPA in prediction of college GPA, especially because GPA is an atheoretical composite that involves not only cognitive skills but also motivation and conscientiousness Dai and Sternberg 2004).

Studying differences among groups can lead to mistaken conclusions. In the Rainbow Project, my colleagues and I sought to create assessments that would reduce ethnic group differences in test scores. Many explanations have been offered for socially defined racial group differences in cognitive-test scores, and for predictive differences for varied ethnic and other groups (Hunt and Carlson 2007; Sternberg et al. 2005).

What did the project find? First, the Rainbow tests shrank ethnic-group differences in comparison with traditional tests of cognitive skills like the SAT. Second, more specifically, Latino students benefited the most from the mitigation of group differences. African-American students, as well, seemed to show a reduced difference from the European-American (white) mean for most of the Rainbow assessments, although a nontrivial difference remained on the practical performance measures. We unfortunately were unable to gain specific socioeconomic

information for individual students or groups of students, desirable though it would have been to obtain such information.

To be more precise, omega-squared, the measure we used to compare white and Asian-American subjects with underrepresented minorities (African-American, Latino-American, American-Indian), were .09 and .04 for the SAT Reading and SAT Mathematics tests, respectively. For the Rainbow tests, omega-squared ranged from .00 to .03 with a median of .02. Thus, our tests showed reduced ethnic-group differences relative to the SAT.

Although the group differences were not eliminated, the results show that assessments can be created that lessen ethnic and racial group differences on college-admissions assessments, particularly for historically disadvantaged groups like African-American and Latino students. Thus, it is possible to reduce adverse impact in undergraduate admissions.

The Rainbow assessments essentially doubled prediction of first-year college GPA in comparison with the SAT alone. Moreover, the Rainbow assessments add prediction substantially beyond the contributions of the SAT and high school GPA.

Although Rainbow and the related projects described below are intended to predict success in the university and beyond, often the curricula of elementary and secondary schools and universities do not develop the creative, practical, and wisdom-based skills needed for adaptation to the circumstances of people's lives. A model for integrating creative, analytical, practical, and wisdom-based skills into the curriculum is ACCEL, or Active and Concerned Citizenship and Ethical Leadership (Sternberg 2016, 2017). The curriculum emphasizes the synthesis of the analytical skills that schools already teach with the additional skills that are important for life success.

Related Projects

The principles behind the Rainbow Project apply at other levels of admissions as well. For example, Jennifer Hedlund and her collaborators (Hedlund et al. 2006) showed that the ideas of the theory of successful intelligence also could be applied to admission to business schools. The goal this project, the University of Michigan Business School Project, was to show if it was possible to improve prediction of success in business beyond that provided by a standardized test. The focus of the project was on practical intelligence. Students were given either long or short scenarios from which they were asked to make situational judgments. The scenarios measured practical reasoning in various domains of business success, such as management, marketing, technology, and sales. The result was an increase in prediction and a decrease in ethnic- (as well as gender-) group differences. Moreover, the test predicted results on an important independent project that were not predicted by the GMAT (Graduate Management Admission Test). In other words, the test successfully supplemented the GMAT in predicting success in an MBA program.

In another project, the goal was to determine whether supplementing difficult tests used for college admissions and placement could increase content validity—the extent to which tests actually covered the full content needed to understand a course—and also decrease ethnic-group differences relative to a conventional test. Steven Stemler and colleagues found that including creative and practical items in augmented physics, psychology, and statistics AP (Advanced Placement) Examinations, in addition to the memory and analytical items already in the AP tests, resulted in more comprehensive coverage of course material, including both creative and practically oriented content as well as analytical content (higher content validity),

and also reduced obtained ethnic-group differences on the tests (Stemler et al. 2006; Stemler et al. 2009).

Would assessments such as those of Rainbow actually work in high-stakes assessment situations? The results of a second project, Project Kaleidoscope, addressed this question.

Rainbow was a low-stakes test. Subjects merely completed a series of test questions in exchange for money or course credit. The problem with such a system is that, for many students, it does not adequately incentivize performance. Results obtained under low-stakes conditions may or may not generalize to higher stakes conditions.

The Kaleidoscope Project

Beginning in 2006 and continuing even to the present day, Tufts has placed on college applications for all of the over 15,000 students applying annually to Arts, Sciences, and Engineering essay-based questions designed to assess the elements of successful intelligence—wisdom, analytical and practical intelligence, and creativity synthesized (Sternberg 2003). The project has been called Kaleidoscope. Kaleidoscope has gone beyond Rainbow to incorporate into its assessment the psychological attribute of wisdom (Sternberg 2009; Sternberg and Coffin 2010).

When this research project commenced, students were not required to do the Kaleidoscope essays. Rather, the essays were strictly optional. For whereas the Rainbow Project was conducted as a separate but experimental high-stakes test administered with a proctor, the Kaleidoscope Project was implemented as an actual section of the Tufts-specific supplement to the Common Application for college admissions. In real-world admissions, it just was not practical to administer an additional high-stakes test. The complete set of items for the initial years of the project can be found in the appendix of Sternberg (2010).

It was not feasible for Kaleidoscope essays to be mandatory. Applicants were encouraged to write just one essay so as not to require too much of them. The goal was not to present to students applying to Tufts an application that would prove burdensome, especially in comparison with the applications of competitors.

According to the theory of successful intelligence, successfully intelligence involves capitalization on strengths and compensation for or correction of weaknesses. By asking students to do just one essay, the applicants could capitalize on a strength.

Two examples of titles on the basis of which students could write creative essays were "The End of MTV" or "Confessions of a Middle-School Bully." A further type of creative question asked applicants what the world would be like if a particular historical event had turned out differently, for example, if the Nazis had won World War II. Still another type creative question provided students with an opportunity to design a new product or create an advertisement for a new product. Students also could design a scientific experiment. An essay encouraging practical thinking asked applicants to say how they had persuaded others of an unpopular idea in which they believed. A wisdom-based essay allowed students to write about how a passion they experienced in high school later could be turned toward achieving a common good.

Tufts assessed quality of creative and practical thinking in the same way as in the Rainbow Project. The university assessed quality of analytical thinking by the organization, logic, and balance of the essay. It assessed wise thinking by the extent to which an essay represented the seeking of a common good by balancing one's own, others', and institutional interests over the long as well as the short term through the use of positive ethical values.

The goal in Kaleidoscope was not to replace the SAT, ACT, or other traditional admissions indices such as GPAs and class rank. Instead, the goal was to re-conceptualize applicants in a broader way—in terms of their academic/analytical, creative, practical, and wisdom-based thinking skills. The researchers used the essays as one but not as the sole source of information. For example, some students submitted creative work in a portfolio, and this work also could be counted in the creativity rating. Evidence of creativity provided by the receipt of prizes or awards also was deemed to be relevant. Thus, the essays were major sources of information, but other information, when available, was used as well.

Admissions officers evaluated applicants for creative, practical, and wisdom-based skills, if sufficient evidence was available, as well as for academic (analytical) and personal qualities in general.

In the first year of Kaleidoscope, approximately half of the academically qualified applicants for admission completed an optional Kaleidoscope essay. In subsequent years, about two thirds completed a Kaleidoscope essay. Merely writing the Kaleidoscope essays did not improve chances of admissions. However, *quality* of essays or other evidence of creative, practical, or wisdom-based abilities did improve chances. For those applicants rated as an "A" (top rating) by a trained admission officer in any of these three categories, average rates of acceptance were roughly double those for applicants not receiving an A. Because of the large number of essays per year (over 8000), only one rater rated applicants except for a small sample used to ensure that inter-rater reliability was sufficient, which it was.

Kaleidoscope was an "action-research" project, in the sense that it was a research project but also was an actual part of the Tufts admissions process. In pilot investigations, the inter-rater reliability of the Kaleidoscope ratings was at the .8 level, but because of the 15,000+ applicants and the small professional admissions staff, we were able to have only one trained rater for Kaleidoscope ratings used for Tufts admissions.

Sometimes new kinds of assessments are introduced that do not look like conventional standardized tests but that actually measure the same skills as are measured by the conventional tests. Convergent-discriminant validation is needed for assessments: Would assessments correlate with other measures with which they should be correlated and would they *not* correlate with other measures with which they should not correlate? The correlations of the assessments with an overall academic rating taking into account SAT scores and high school GPA were relatively low but statistically significant for creative, practical thinking, and wise thinking. The correlations of the assessments with a rating of quality of extracurricular participation and leadership were higher and moderate for creative, practical, and wise thinking. Thus, the pattern of convergent-discriminant validation was what the investigators had sought.

In the first year of Kaleidoscope, the academic credentials (SATs and GPAs) of applicants to Arts and Sciences at Tufts rose slightly. Kaleidoscope success predicted success in first-year GPA. Moreover, Tufts had substantially lower numbers of students in what before had been the bottom third of the pool in terms of academic quality. Some number of those students, seeing the new application, apparently decided not to apply to Tufts. In contrast, many more highly qualified applicants sought admission. The researcher left Tufts after the first-year data were analyzed so could not speak to data from later years, but the procedure is still used, indicating that at least the university views it as a success.

A fear of some faculty and administrators was that use of Kaleidoscope would lower the academic quality of the student body. In fact, the opposite happened. Instead, the applicants who were admitted were *more highly* qualified, and in a broader way. Moreover, the subjective

responses of applicants and their parents were very positive. Applicants especially like an application that enabled them better to show who they are.

Tufts did not get meaningful statistical differences in scores across ethnic groups. This result was in contrast to the results for Rainbow, which reduced but did not eliminate ethnic-group differences. After a number of years during which numbers of applications from underrepresented minorities remained relatively constant, Kaleidoscope seemed to produce an increase (although real-world college admissions are complex and it is difficult to know with any certainty what causes what). In the first year, applications from African-Americans and Latino Americans increased significantly, and admissions of African-Americans increased 30% while admissions of Latino Americans increased 15%. There is no proof that Kaleidoscope was causal of this difference. However, the fact that Kaleidoscope did not show ethnic-group differences whereas standardized tests did is at least suggestive of some kind of important role of Kaleidoscope in the difference: In previous years, many of the minority students with lower SATs did not have sufficient counterbalancing information in their admissions folder to be admitted.

These results, like those from the Rainbow Project, demonstrated that colleges can increase academic quality and diversity simultaneously. Moreover, they can so for an entire college class at a major university, not just for small samples of students at some scattered schools. Kaleidoscope also let students, parents, high school guidance counselors, and others know that there is a more to a person than the narrow spectrum of skills assessed by standardized tests; moreover, these broader skills can be assessed in a quantifiable way.

The projects described above use the theory of successful intelligence to construct tests that can be used for various kinds of admissions or diagnostic purposes. But they were not designed to be used to measure adaptive skills of individuals with highly varying cultural backgrounds. How would one measure such adaptive skills?

Measuring Adaptive Skills

Intelligence tests are designed to predict adaptation to one's environment. As Flynn (2016) has pointed out, the mean and range of IQ are not constant. Rather, increasing (or decreasing) IQs across generations reflect changing demands in terms of adaptation. Today, technological skills are at a premium, and the abstract analytical skills measured by IQ tests are well tuned to such an environment. So as the environment requires increasingly complex technological skills, IQs are likely to increase. But in an increasingly technologically deskilled environment, IQs likely will go down while other skills more tuned to those environments (e.g., hunting or foraging skills) may increase. It is not that IQ is irrelevant to hunting: figuring out what the prey is likely to do may well be predicted by IQ tests. But high IQ will be no substitute for an accurate and steady aim and an understanding of the prey's habit and escape strategies. Knowing where an animal is will be no help if one cannot incapacitate the animal, whether by a rifle, bow and arrow, or whatever. Many "high IQ" individuals would be helpless in such environments. Much of the emphasis of the work in this area has been in the understanding and measurement of practical skills, as characterized by the theory of successful intelligence.

In the field of cultural studies of intelligence, progress has been made, largely due to the pioneering work of Luria (1976). Luria, in testing individuals in non-European cultures, found that the problems that were alleged to measure intelligence in European populations did not well do so in other cultures, because the individuals did not accept the presuppositions of the

problems they were given. For example, when Uzbekistan peasants were given a syllogisms problem, such as, "There are no camels in Germany. The city of B. is in Germany. Are there camels there or not?" subjects could repeat the problem precisely and then answer "I don't know. I've never seen German villages…" The subjects did not accept the problems in the abstract modality in which they were presented. Of course, one could argue that they could not do so. There are alternative interpretations of these findings (Cole et al. 1971). But then, Cole et al. (1971) found that Kpelle people are experts in measuring rice. Most of us would not even think about measuring rice, much less how to do it better or worse. The point is that people acquire expertise—mathematical, linguistic, or otherwise—in terms of the sociocultural contexts in which they are socialized.

A Study in Rural Kenya The extent to which practical adaptive competencies can be hidden from us as psychologists is demonstrated in research among Luo children in rural Kenya (Sternberg et al. 2001).

Consider a child in a small rural Kenyan village. The senior author of the study first learned something of these children in a discussed with a parasitologist, then at Oxford. The parasitologist, Kate Nokes, mentioned that children in rural villages in Kenya would know the names of 80, 90, or even 100 natural herbal medicines that could be used to combat parasitic illnesses. Such knowledge is extremely relevant for adaptation by these children because parasitic illnesses are endemic in the regions in which they live and interfere greatly with the children's ability to function, to the point that children may have to stay home from school or work because they are too ill to be effective in school or on the job.

If knowledge of natural herbal medicines was just a proxy for general ability (*g*) or academic knowledge, then a teacher might predict the children's knowledge from conventional tests, standardized or otherwise. But suppose that such knowledge was not predictable from conventional tests, then knowing something of children's' ability to learn, as evidenced by their knowledge of natural herbal medicines, might be useful information for a teacher to have in assessing which children could be more successful in learning tasks than perhaps they appeared to be on the basis of their school work.

The child's prospects in some of these rural Kenyan villages are rather limited. Schooling beyond the early years is considered largely a waste of time because there is little need for academic skills in the village. But, there is a need for a knowledge of natural herbal medicines that can be used to treat that various parasitic illnesses prevalent in the region, such as malaria, schistosomiasis, hookworm, whipworm, and the like. Consider a problem presented to subjects:

- "A small child in your family has homa. She has a sore throat, headache, and fever. She has been sick for 3 days. Which of the following five Yadh nyaluo (Luo herbal medicines) can treat homa?

   i.   Chamama. Take the leaf and fito (sniff medicine up the nose to sneeze out illness).*
   ii. Kaladali. Take the leaves, drink, and fito.*
   iii. Obuo. Take the leaves and fito.*
   iv. Ogaka. Take the roots, pound, and drink.
   v.   Ahundo. Take the leaves and fito."

There are multiple correct answers, which are asterisked. Once again, no one would expect a typical US college professor or student to be able to answer such questions at better than a chance level. Why should they? The knowledge probably has no real adaptive value for them (unless they are studying cultural psychology or anthropology). But for children growing up in an environment where the major threat to adaptive success is parasitic illness, such knowledge is extremely important.

In terms of the theory of successful intelligence, this problem would be a practical one, in that it is used for adaptation to everyday life, and indeed, can be a matter of life or death in the case of serious parasitic illnesses. Learning about this practical situation activates the three knowledge-acquisition components described above. The children have to decide what information in the problem is relevant—for example, the name of the disease and the symptoms. The children also have to selectively compare the information in the problem to what they already know. What are possible treatments for homa? Finally, the children have to combine the information in the problem regarding homa with their prior knowledge about homa to choose the best of the answer options presented (the performance component of justification). The processes are comparable to those in the socks problem, but the homa problem is practical to the children's everyday lives, whereas the socks problem—figuring out how many socks have to be withdrawn from a drawer to have two of the same color—is more abstract and unlikely to be one that they would be inclined to solve in their everyday lives.

The investigators also tested the children for their vocabulary levels in Dholuo, their home language, and in English. Such measures assess the so-called crystallized intelligence. They also used geometric matrix problems to measure their so-called fluid intelligence. The expectation, based on work that had been done on practical intelligence (Sternberg et al. 2000), was that the knowledge of the natural herbal medicines would show at most a weak positive correlation with scores on the standardized ability tests.

To the investigators' surprise, there were significant correlations, but they were negative. For example, the correlation with vocabulary (English and Dholuo combined) was −.31 ($p < .01$). This left the investigators, at first, puzzled and might leave other psychologists puzzled as well because tests of fluid and crystallized abilities typically show a positive manifold, that is, a pattern of positive correlations throughout that yields a general factor (g) when the tests are factor-analyzed. But there is a logic to the negative pattern of correlations. What the correlations showed is the extent to which patterns of relationships among assessments may be influenced not only by characteristics internal to individuals but also to the environmental contexts in which they live.

In particular, in these villages, the anthropologists on the Kenya research team (Prince and Geissler) found that students who were viewed as adaptively and practically competent by the adults in the villages would be selected by certain adults to do apprenticeships with them (see Prince et al. 2001). Such apprenticeships would take them out of formal schooling (see also No Swots please 2002). Whereas many of us greatly value formal schooling, such schooling is less valued among the Luo because, in the village, it does not ultimately lead to gainful employment, perhaps except as a teacher. But because most children will leave formal schooling in elementary school, there is a not a great need for teachers in the Western sense. Rather, there is a need for mentors who will apprentice children to learn the skills that can lead the children to earn a living. So, perhaps oddly by our way of thinking, the children viewed as adaptively competent are whisked out of school whereas the children not viewed as being quite so competent are left in school, where they continue to acquire the knowledge and skills measured by conventional standardized tests but not the knowledge and skills that will earn

them a living. As a result, children who acquire more formal knowledge in turn acquire fewer less of what might be called "practical knowledge," and hence fewer adaptive competencies.

Another way of viewing what happened in the study in Kenya is in terms of cultural capital (Bourdieu 1986; Byun et al. 2012)—the social assets that produce advancement and mobility within a culture. The Kenyan children had a great deal of cultural capital, but not of the kind valued by Western society and its standardized tests. So locally, the children who were skilled in, for example, recognizing parasitic illnesses were viewed as having high cultural capital, but in terms of knowledge and skills recognized as important by Western culture, they were viewed as lacking important cultural capital.

One might be inclined to think that the phenomenon observed in Kenya is limited to cultures remote from ours, but that really is not the case. In our culture as well, gaining more education can lead to reduced societally valued outcomes, such as money. For example, students with a two-year MBA generally will earn substantially more money than students with a PhD earned over four, five, six, or more years. In Silicon Valley, the entrepreneurs who run start-up companies often are individuals who have nothing more than a bachelor's degree, if that; they hire PhDs to work for them, at salaries considerably lower than their own. The grade level at which additional formal schooling leads to certain reduced societally valued outcomes is different but the principle is the same: at some point, additional schooling and acquisition of associated academic knowledge and skills may lead to a reduction rather than an increase in certain societally valued outcomes. This is even more the case in most other countries of the world, where college and university professors are paid far less than they are in the USA. German universities, for example, generally pay less than American universities, and the national pay scale for professors recently was reduced.

None of this is to say that everyone living in a developing country wants to stay "developing," at least in terms of Western notions about what "developing" means. In Africa and elsewhere, governments are investing large amounts of money in Western types of schooling to help their children prepare for a globalized society. The point here is not that this is "bad" in any sense. Rather, the point is that the children who are being prepared for globalized society may have developed important indigenous skills that can be leveraged for Western education, but that are not recognized or measured by conventional tests of abilities and achievement.

All that said, according to the theory of successful intelligence, criteria of success are individually determined: people decide what is important to them personally. So money may be important to some people, and of only the slightest importance or even of no importance to others.

A Study Among Yupi'k Children In work with Alaskan Yup'ik schoolchildren (Grigorenko, Meier, Lipka, Mohatt, Yanez, & Sternberg 2004), for example, we discovered that the Native American children were able to navigate on a dog sled from one distant village to another across what to us (and probably you) would have seemed to be a perceptually uniform field of vision. If you (or the children's non-Native American schoolteachers) attempted to go from one village to another on such a dog sled, you probably would get lost in the wilderness and die. Signals for navigation are there; so most of us just would not see them. Similarly, the Puluwat can navigate across long distances in the sea under circumstances in which meaningful signals also would elude us (Gladwin 1970).This is not to say that the Eskimos' skills are somehow superior to the teachers'. Rather, adaptive skills depend on the environment in which one is raised. When the academic establishment values children's skills

in terms of what it thinks is important, it may neglect the skills that children have adapted in their home environments that could be leveraged toward success in other environments.

The importance of context is shown by the kinds of practical knowledge that children develop in order to adapt to their environments. Consider two examples.

Imagine living in a hunting-gathering society. Many Yup'ik Eskimos in Alaska live in such a society, where hunting and gathering are joined by fishing as means of putting food on the table. The knowledge and skills you need to survive in such an environmental context are rather different from those of, say, an individual who has spent his life as a professor. The professor (or college student, for that matter) might do well on an SAT question or on a question about what or how to order in a restaurant. He or she might not fare as well on a question developed for assessing Yup'ik children.

"When Eddie runs to collect the ptarmigan that he's just shot, he notices that its front pouch (balloon) is full of ptarmigan food. This is a sign that:

- there's a storm on the way.*
- winter is almost over.
- it's hard to find food this season.
- it hasn't snowed in a long time."

The correct answer is asterisked. Of course, there is no reason why the typical college student or professor would need to know the answer to the question about the ptarmigan. But similarly, it is unclear that the Alaskan Yup'ik student would need to do well on the SAT or restaurant question if he or she plans to remain in a coastal Yup'ik village with no restaurants and no need to read complicated texts or perform complex mathematical operations. The knowledge that is useful depends on the context.

Once again, the child solving such a practical problem uses selective encoding to decide what information in the problem is relevant (the front pouch is full of ptarmigan food), selective comparison to retrieve from long-term memory the meaning of a full front pouch, and selective combination is used to put together the fullness of the front pouch with knowledge about what ptarmigan do when a storm is about to approach. The performance component of justification is used to select the best answer. Here, as with the "homa" problem, the knowledge-acquisition components applied are the same as would be applied in a more academic problem—what differs is the practical context of the ptarmigan problem.

The investigators found that urban students (from Dillingham, a city in Alaska that, although small by the standards of most states would count as fairly large in Alaska) outperformed rural students on conventional tests of fluid and crystallized abilities, but that the Yup'ik Eskimo children outperformed the urban children on tests of knowledge of adaptive competencies relevant to the Yup'ik environment. Moreover, tests of practical knowledge predicted hunting skills whereas conventional standardized tests did not. These results, like the ones with the Luo children in Kenya, also could be viewed in terms of different kinds of cultural capital being valued in different settings (Bourdieu 1986; Byun et al. 2012).

A Further Study Among Yup'ik Eskimo Children At this point, the investigators wondered whether the practical knowledge and adaptive competencies that Yup'ik and other children have for knowledge not learned in school might be leveraged to help them perform better in school. In other words, might the children do better in the acquisition of academic

knowledge and skills if teachers enabled the children to utilize their practical knowledge in the context of the classroom?

In a further study (Sternberg et al. 2007), we taught Yup'ik children the plane-geometry concepts of perimeter and area using either textbook presentation or a novel presentation prominently featuring fish racks, which are an integral part of the environment of children in rural fishing villages. We found that the Yup'ik children who were taught via the fish racks outperformed the children who were taught using conventional textbook presentation. So the children were good problem solvers in a traditional sense (see Davidson and Sternberg 2003), but that good problem solving needed familiar ecological content to elicit it.

Perhaps the finding is not altogether surprising. Many of us who are parents find that our children learn better when they are taught in ways that capitalize on interests of even passions they may have, whether for handheld phones, computers, art, music, or whatever. In this respect, the Yup'ik children are no different from our own: they learn better when they can relate in a meaningful way to what they are learning.

Measures of adaptive skills provide ways to assess strengths that conventional tests often hide. The processes of solution are the same as those measured by more academic tests—what differs is the context. The children are solving problems for which they have an adequate real-life context and, further, that they are motivated to solve because the problems are life-relevant.

The instructional study among the Yup'ik deals with an issue that all schools should confront: is what they are teaching and how they are teaching it a good match to the students they are teaching? Are the schools teaching the knowledge and skills that students will need to cope in the lives they will live? On the one hand, there are certain core skills and packets of knowledge that virtually all students will need to succeed in their lives (e.g., how to read, how to add, etc.). On the other hand, these skills and packets of knowledge, if taught in a way that makes sense to the students in the context of their lives, may be learned and retained much better. In life, all students will need to cope with novel tasks and situations (creative skills), analyze what they see and hear to determine if it makes sense (analytical skills), transfer what they learn in school to their everyday lives (practical skills), and seek a common good beyond just what is best for them (wisdom-based skills). Hence, I believe, it makes sense to teach and test in ways that transmit and assess these skills that are life-relevant for everyone. Moreover, if one teaches not only about the challenges one encounters in one's own sociocultural context but also about the challenges others confront, students are more likely to appreciate the challenges others face and perhaps even apply more diverse perspectives to their own problems.

Another way of assessing such strengths is through dynamic assessment.

Dynamic Assessment

All of the studies described above have used conventional static assessment—the examiner presents a series of test items, the subjects take the tests, the subjects complete the tests, the tests are scored, and subjects may or may not receive their scores. But a different procedure of testing also may be useful for populations that are not accustomed to standardized testing (Grigorenko and Sternberg 1998; Sternberg and Grigorenko 2002). This procedure is dynamic testing, as described below.

In a study in Tanzania (Sternberg et al. 2002), we investigated whether a form of testing called "dynamic assessment" (Sternberg and Grigorenko 2001, 2002) might better capture the strengths of children in developing countries, especially those with parasitic or other illnesses.

When we think of testing, we typically think in terms of conventional "static" testing. Basically, someone administers a test; students or others respond to the questions on the test; then the examinees get a score indicating how well they performed. The test is "static" in that it is a snapshot at one period of time—it is essentially frozen in time. Dynamic testing is different in that it examines performance over a period of time. That is, it looks at how test performance evolves from one time-period to another. There are two common kinds of dynamic testing, what can be referred to as "sandwich" and "layer cake" models of testing.

In sandwich testing, the examinee takes a pretest. Then he or she learns something. Then the examinee is tested again. One is thereby able to assess learning that occurred at time of test. In layer-cake testing, the child is given testing items, usually difficult ones, that she typically cannot initially solve. Then, she is given a series of prearranged clues that will help her solve the problem. The dependent variable of greatest interest is the number of clues she needs in order to solve each problem. Thus, there is not just one overall learning experience, but layers of learning experiences that culminate either in solution to the problem or failure to solve it, despite the clues. Both kinds of dynamic testing are based on the work of Vygotsky (1980) who proposed that individuals have a zone of proximal development in which they readily can learn new knowledge and skills. Knowledge and skills outside this zone is learned only with great difficulty, if at all. Feuerstein (1979) capitalized on this concept to create a dynamic test of learning potential, the Learning Potential Assessment Device (LPAD).

The reason such testing can be important for children in the developing world is that the knowledge base with which they come into a test may be very different from the knowledge base with which a child in the developed world, and especially those living in middle- or upper middle class socioeconomic contexts, enter the testing situation. The result can be that the children from the developing world never seem to have, on average, quite the knowledge and academic skills of children from the developed world. The advantage of a dynamic test is that it can capture learning at time of test, so that the tester better (although not completely) controls for the background knowledge and skills with which the children approached the test.

In the Tanzania study, children were given either a static or dynamic version of three different tasks: syllogisms, sorting, and twenty questions. In the static test, they simply received a pretest and a posttest that was essentially an alternate form of the pretest. In the treatment group, children received instruction sandwiched in between the pretest and posttest. On all three tasks, dynamically tested children improved significantly from the pretest to the posttest. In other words, the dynamic test did enable the children to learn from experience. More important, the correlation between pretest and posttest was about .8 in the control group but .3 in the experimental group. In other words, the instructional treatment changed the rank orders of the students on the posttest relative to the pretest. In the experimental group, posttest scores correlated significantly more highly with other cognitive tests than did the pretest measures. In other words, the posttest—the test administered after dynamic instruction— proved to be the better predictor than the pretest of how well children would succeed in a variety of cognitive-testing situations. This may be because when these children go into a testing situation, they just do not have the prior familiarity with it that many children in the developed world would have. The instruction enables them to have an orientation to the kind of test that they will take so that they are more prepared on the posttest to show what they really can do.

In the study described above, dynamic testing was done in a single session. The question might arise of what would happen if the children were tested over multiple sessions, giving them even more time to familiarize themselves with the kinds of questions on which they will be tested. Although this may sound like a kind of testing we would never use in the West, it actually is similar to the situation of students in the West who take a standardized test but previously have taken a course or used a book to help prepare them for the standardized test. They take a practice pretest, then get instruction, then take a practice test, then get instruction, then take a practice test, and onward until they believe they are ready for the actual test (Grigorenko et al. 2006).

Children in a control group (uninfected) improved more over time in their cognitive skills than did infected (experimental-group) children in either a medically treated or placebo-treated condition. But the medically treated children improved more than did the placebo-treated ones. In other words, anti-parasitic medicines, such as albendazole, not only improve children's health—they also help improve their cognitive performance. From the standpoint of cognitive performance, it is better not to have the parasitic illness in the first place, but if one does, it is better to be treated with an anti-parasitic drug than with a placebo.

Note that the kind of dynamic testing we did is not merely one of giving children practice in particular kinds of items. Many of the children initially do not even understand the items or what they are expected to do with them. They need instruction in order to understand what solution of the items entails.

In sum, dynamic testing can be useful in the armamentarium of techniques used to assess children who do not come from mainstream cultures. Where does this technique and all the other techniques leave us?

Conclusion

In this article, it has been suggested that conventional standardized testing may be most appropriate for members of groups that who have grown up in an environment that highly values the use of abstract analytical skills and who are accustomed to taking standardized tests. The tests may be less appropriate for those whose environments have emphasized other skills that are adaptive in those environments. Because intelligence is a matter of adaptive skills, the testing of intelligence should reflect, at least in large part, the adaptive skills into which individuals have been socialized in order to succeed in their environments. These skills in turn later may be transferred to other environments.

The four interrelated approaches reviewed are (a) use of broader measures, (b) performance- and project-based assessments, (c) direct measurement of knowledge and skills relevant to environmental adaptation, and (d) dynamic assessment. The advantage of the first approach is that it can capture knowledge and skills that are missed by assessments that focus on conventional academic knowledge and skills. The advantage of the second approach is that it goes beyond short answer and multiple choice, enabling individuals to show what they can do in constructing products (see also Gardner 2011). In our own research, we found that multiple-choice tests, regardless of whether they were supposed to measure analytical, creative, or practical skills, all loaded highly on a general-intelligence (g) factor. The advantage of the third approach is that it measures skills directly relevant to the environmental adaptive challenges that individuals face, which standardized tests often do not do, especially for nonstandard populations different from those for which the tests were originally intended.

And the advantage of the fourth approach is that it enables individuals to learn in the testing situation and, thus, provides the test-takers with familiarity with types of assessments that some of them who have not had typical Western education may never have encountered before.

One could argue that conventional tests are not designed to measure adaptation to any environment, but rather, to environments in which Western schooling dominates. This is not how the tests have been used, however. They have been widely translated and applied in an extremely broad range of environments, in some of which they may have been more appropriate than others (Sternberg 2004). Even if one is interested only in adaptation to industrialized Western environments, one might suggest that if the tests are sold as tests of cognitive skills or of intelligence, then they need to match the cognitive skills that have been developed in the individuals' rearing environments. Those are the skills that can be transferred, not skills that the individuals have never developed, at least in the context of abstract analytical reasoning.

When conventional tests are used, at the very least, techniques can and should be used to minimize stereotype threat. These techniques include measures such as self-affirmation and telling students that concerns about social belonging tend to lessen over time (Cohen et al. 2006; Walton and Cohen 2003). Another technique is to emphasize to students the malleability of intelligence (Aronson et al. 2002; Good et al. 2003).

In the Kaleidoscope project at Tufts University, investigators found that using broader tests, for high stakes, resulted in the admission of students who never would have been admitted otherwise (Sternberg 2010, 2016). So, the tests can make a difference. Moreover, in the overwhelming majority of cases, these students succeeded in the college environments. Investigators have not used dynamic tests in these environments, but had they done so, they might have uncovered even more talent.

Others, of course, have been concerned with many of the same issues. For example, the Coalition of Essential Schools (http://essentialschools.org/common-principles/) has proposed a set of common principles that overlap substantially with many of the recommendations in this article. These principles include, for example, (a) emphasizing depth over coverage so that students can learn to think well rather than just memorizing a lot of unrelated facts; (b) personalization, so that students are treated as individuals rather than as members of groups that may be largely irrelevant to their particular needs; (c) resources that are dedicated to teaching and learning, so that the focus is on students rather than on administrative or overhead costs; and (d) a tone of decency and trust, so that students develop the wisdom-based and ethical skills that are so important to success in life, almost without regard to how "success," or at least prosocial success, is defined.

There is a lot of talent out there waiting to be recognized. We have at least some preliminary tools to recognize this talent, although we also have a long way to go. Do we have the foresight to use what we have and to develop the new tools we still need? And do we have the insight to develop the school curricula to help *all* children optimally to develop their talents, regardless of the sociocultural context in which they grew up?

So, where do we end up? The article leads to five conclusions.

First, if our schools are to develop active concerned citizens and ethical leaders, they need to develop and assess not only knowledge and analytical skills, but also creative, practical, and wisdom-based skills. Students develop analytical skills by analyzing, comparing and contrast-ing, evaluating, and critiquing. They develop creative skills by creating, inventing, discover-ing, imagining, and exploring. They develop practical skills by implementing, applying,

practicing, using, and persuading. And they develop wisdom-based skills by using their knowledge and creative, analytical, and practical skills to seek a common good, over the long- as well as the short-term period, through the infusion of positive ethical values.

Second, schools need to help students capitalize on their own individual strengths and to compensate for or correct their own individual weaknesses. No one is good at everything. This article has pointed out a wide array of strengths students can have. Instruction and assessment should be oriented, in part, toward identifying these strengths—which go well beyond the boundaries of conventional standardized testing—and also toward helping students compensate for and correct weaknesses.

Third, schools need, at least in part, to instruct and assess in ways that takes into account the cultural capital that students bring to instructional and assessment situations. Students from nontraditional backgrounds may have both a broad range and a deep development of skills, but they not be skills that traditionally make their way into Western school curricula. Students can learn better if they are taught and assessed in ways that make at least some contact with the kind of upbringing they have had.

Fourth, students may, under some circumstances, benefit from dynamic forms of assessment, which enable them to become familiar with kinds of testing that they are not familiar with before they encounter a test. There is evidence that dynamic testing can benefit students under a variety of circumstances (Sternberg and Grigorenko 2001, 2002).

Finally, context-sensitive cognitive testing is based on the notion that abilities always are expressed in some kind of context. The Flynn effect has shown that even items that once were thought to be culture fair or even culture free are not. All testing, including standardized testing, is in some kind of context. We can never test in a context-free or fully context-fair way. But, if we, as educators, are sensitive to context, at least we can give children the chance to demonstrate their strengths rather than merely to showcase their weaknesses.

References

Abrams, L. S., & Terry, D. L. (2014). "You can run but you can't hide": How formerly incarcerated young men navigate neighborhood risks. *Children and Youth Services Review, 47*, 61–69.

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*, 113–125.

Arteche, A., Chamorro-Premuzic, T., Ackerman, P., & Furnham, A. (2009). Typical intellectual engagement as a byproduct of openness, learning approaches, and self-assessed intelligence. *Educational Psychology: An International Journal of Experimental Educational Psychology, 29*, 357–367.

Berry, J. W. (1974). Radical cultural relativism and the concept of intelligence. In J. W. Berry & P. R. Dasen (Eds.), *Culture and cognition: readings in cross-cultural psychology* (pp. 225–229). London: Methuen.

Berry, J. W., & Irvine, S. H. (1986). Bricolage: savages do it daily. In R. J. Sternberg & R. K. Wagner (Eds.), *Practical intelligence: nature and origins of competence in the everyday world* (pp. 271–306). New York: Cambridge University Press.

Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (1992). *Cross-cultural psychology: research and applications*. New York: Cambridge University Press.

Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 46–58). Greenwich: Greenwood.

Byun, S., Schofer, E., & Kim, K. (2012). Revisiting the role of cultural capital in East Asian educational systems: the case of South Korea. *Sociology of Education, 85*(3), 219–239.

Cattell, R. (1949). *Culture free intelligence test, scale 1, handbook*. Champaign: Institute of Personality and Ability Testing.

Cattell, R. B., Krug, S. E., & Barton, K. (1973). *Technical supplement for the culture fair intelligence tests, scales 2 and 3*. Champaign: Institute for Personality and Ability Testing.

Ceci, S., & Roazzi, A. (1994). The effects of context on cognition: postcards from Brazil. In R. J. Sternberg & R. K. Wagner (Eds.), *Mind in context: interactionist perspective on human intelligence* (pp. 74–101). New York: Cambridge University Press.

Cole, M. (2017). Idiocultural design as a tool of cultural psychology. *Perspectives on Psychological Science.* http://journals.sagepub.com/doi/full/10.1177/1745691617708623.

Cole, M., Gay, J., Glick, J. A., & Sharp, D. W. (1971). *The cultural context of learning and thinking: an exploration in experimental anthropology.* New York: Basic.

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: a social-psychological intervention. *Science, 313*(5791), 1307–1310.

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

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harpercollins.

Dai, D. Y., & Sternberg, R. J. (Eds.). (2004). *Motivation, emotion, and cognition: integrative perspectives on intellectual functioning and development.* Mahwah: Lawrence Erlbaum Associates, Inc. Publishers.

Davidson, J. E., & Sternberg, R. J. (Eds.). (2003). *The psychology of problem solving.* New York: Cambridge University Press.

DeYoung, C. G. (2011). Intelligence and personality. In R. J. Sternberg & S. B. Kaufman (Eds.), *Cambridge handbook of intelligence* (pp. 711–737). New York: Cambridge University Press.

Feuerstein, R. (1979). *The dynamic assessment of retarded performers: the learning potential assessment device, theory, instruments, and technique.* Baltimore: University Park Press.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: what IQ tests really measure. *Psychological Bulletin, 101*(2), 171–191.

Flynn, J. R. (2016). *Does your family make you smarter? Nature, nurture, and human autonomy.* New York: Cambridge University Press.

Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or *g*? The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science, 15*(6), 373–378.

Gardner, H. (2011). *Frames of mind: the theory of multiple intelligences.* New York: Basic Books.

Gardner, H., Krechevsky, M., Sternberg, R. J., & Okagaki, L. (1994). Intelligence in context: enhancing students' practical intelligence for school. In K. McGilly (Ed.), *Classroom lessons: integrating cognitive theory and classroom practice* (pp. 105–127). Cambridge: MIT Press.

Gauvain, M. (2013). Sociocultural contexts of development. In P. D. Zelazo (Ed.), *Oxford handbook of developmental psychology (Vol. 2. Self and other)* (pp. 425–451). New York: Oxford University Press.

Gauvain, M., & Munroe, R. L. (2012). Cultural change, human activity, and cognitive development. *Human Development, 55*, 205–228.

Gauvain, M., & Munroe, R. L. (2013). Children's questions in cross-cultural perspective: a four-culture study. *Journal of Cross-Cultural Psychology, 44*, s1148–s1165.

Gladwin, T. (1970). *East is a big bird: navigation and logic on Puluwat Atoll.* Cambridge: Harvard University Press.

Goleman, D. (1995). *Emotional intelligence.* New York: Bantam.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: an intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24*, 645–662.

Greenfield, P. M., Ward, L. M., & Jacobs, J. (1997). You can't take it with you: why ability assessments don't cross cultures. *American Psychologist, 52*, 1115–1124.

Grigorenko, E. L., Geissler, P. W., Prince, R., Okatcha, F., Nokes, C., Kenny, D. A., Bundy, D. A., & Sternberg, R. J. (2001). The organization of Luo conceptions of intelligence: a study of implicit theories in a Kenyan village. *International Journal of Behavioral Development, 25*(4), 367–378.

Grigorenko, E. L., Meier, E., Lipka, J., Mohatt, G., Yanez, E., & Sternberg, R. J. (2004). Academic and practical intelligence: a case study of the Yup'ik in Alaska. *Learning and Individual Differences, 14*, 183–207.

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin, 124*, 75–111.

Grigorenko, E. L., Sternberg, R. J., Jukes, M., Alcock, K., Lambo, J., Ngorosho, D., Nokes, C., & Bundy, D. A. (2006). Effects of antiparasitic treatment on dynamically and statically tested cognitive skills over time. *Journal of Applied Developmental Psychology, 27*(6), 499–526.

Hedlund, J., Wilt, J. M., Nebel, K. R., Ashford, S. J., & Sternberg, R. J. (2006). Assessing practical intelligence in business school admissions: a supplement to the graduate management admissions test. *Learning and Individual Differences, 16*, 101–127.

Hess, T. M., Auman, C., Colcombe, S. J., & Rahhal, T. A. (2003). The impact of stereotype threat on age differences in memory performance. *The Journals of Gerontology: Series B, 58*(1), 3–11.

Hunt, E. (2011). *Human intelligence.* New York: Cambridge University Press.

Hunt, E., & Carlson, J. (2007). Considerations relating to the study of group differences in intelligence. *Perspectives on Psychological Science, 2*, 194–213.

Jensen, A. R. (1998). *The g factor.* Westport: Praeger-Greenwood.

Koenig, K. A., Frey, M. C., & Detterman, D. K. (2008). ACT and general cognitive ability. *Intelligence, 36*, 153–160.

Krechevsky, M. (1998). *Project spectrum: preschool assessment handbook (project zero frameworks for early childhood education, Vol. 3)*. New York: Teachers College Press.

Laboratory of Comparative Human Cognition (1982). Culture and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 642–719).

Lave, J. (1988). *Cognitive in practice: mind, mathematics, and culture in everyday life*. New York: Cambridge University Press.

Lesser, G. S., Fifer, G., & Clark, D. H. (1965). Mental abilities of children from different social class and cultural groups. *Monographs of the Society for Research in Child Development, 30*.

Lubart, T. I., & Sternberg, R. J. (1995). An investment approach to creativity: theory and data. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 269–302). Cambridge: MIT Press.

Luria, A. R. (1976). *Cognitive development: its cultural and social foundations*. Cambridge: Harvard University Press.

Mackintosh, N. J. (2011). *IQ and human intelligence*. New York: Oxford University Press.

Mayer, J. D., & Salovey, P. (1993). The intelligence of emotional intelligence. *Intelligence, 17*, 433–442.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional intelligence: theory, findings, and implications. *Psychological Inquiry, 15*, 197–215.

Mayer, J. D., Salovey, P., Caruso, D. R., & Cherkasskiy, L. (2011). Emotional intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *Cambridge handbook of intelligence* (pp. 528–549). New York: Cambridge University Press.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion, 3*, 97–105.

Nisbett, R. E. (2004). *The geography of thought: how Asians and Westerners thinking differently…and why*. New York: Free Press.

"No Swots please, we're Masai" (2002). *The Economist,* http://www.economist.com/node/1048686.

Nuñes, T., Carraher, D. W., & Schliemann, A. D. (1993). *Street mathematics and school mathematics*. New York: Cambridge University Press.

Ogbu, J. U. (2003). *Black American students in an affluent suburb: a study of academic disengagement*. New York: Routledge.

Ogbu, J. U. (2008). *Minority status, oppositional culture, and schooling*. New York: Rutledge.

Okagaki, L., & Sternberg, R. J. (1993). Parental beliefs and children's school performance. *Child Development, 64*(1), 36–56.

Prince, R. J., Geissler, P. W., Nokes, K., Maende, J. O., Okatcha, F., Grigorenko, E. L., & Sternberg, R. J. (2001). Knowledge of herbal and pharmaceutical medicines among Luo children in western Kenya. *Anthropology & Medicine, 8*(2/3), 211–235.

Raven, J., Raven, J. C., & Court, J.H. (2003) updated 2004. *Manual for Raven's Progressive Matrices and Vocabulary Scales.* San Antonio: Harcourt Assessment.

Rogoff, B. (1991). *Apprenticeship in thinking: cognitive development in social context*. New York: Oxford University Press.

Rogoff, B. (2003). *The cultural nature of human development*. New York: Oxford University Press.

Rogoff, B., Coppens, A. D., Alcala, L., Aceves-Azuara, I., Ruvalcaba, O., Lopez, A., & Dayton, A. (2017). Noticing learners' strengths through cultural research. *Perspectives on Psychological Science,* http://journals.sagepub.com/doi/full/10.1177/1745691617718355.

Sarason, S. B., & Doris, J. (1979). *Educational handicap, public policy, and social history*. New York: Free Press.

Saxe, G. B. (1991). *Culture and cognitive development: studies in mathematical understanding*. Hillsdale: Erlbaum.

Saxe, G. B. (2012). *Cultural development of mathematical ideas: Papua New Guinea studies*. New York: Cambridge University Press.

Schaler, J. A. (2006). *Howard Gardner under fire: the rebel psychologist faces his critics*. Chicago: Open Court.

Serpell, R. (2000). Intelligence and culture. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 549–577). New York: Cambridge University Press.

Serpell, R. (in press). Cultural research insights for cognitive, educational and social psychology. *Perspectives on Psychological Science.*

Shearer, C. B. (2004). Multiple intelligences theory after 20 years. *Teachers College Record, 106*(1), 147–162.

Spearman, C. (1927). *The abilities of man*. New York: Macmillan.

Steele, C. M. (1997). A threat in the air: how stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613–629.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797–811.

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*(2), 344–376.

Stemler, S., Sternberg, R. J., Grigorenko, E. L., Jarvin, L., & Sharpes, D. 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. (1983). Components of human intelligence. *Cognition, 15*, 1–48.

Sternberg, R. J. (1984). What should intelligence tests test? Implications of a triarchic theory of intelligence for intelligence testing. *Educational Researcher, 13*, 5–15.

Sternberg, R. J. (1985a). *Beyond IQ: a triarchic theory of human intelligence*. New York: Cambridge University Press.

Sternberg, R. J. (Ed.). (1985b). *Human abilities: an information-processing approach*. San Francisco: Freeman.

Sternberg, R. J. (1985c). Teaching critical thinking, part 1: are we making critical mistakes? *Phi Delta Kappan, 67*, 194–198.

Sternberg, R. J. (Ed.). (1988a). *Advances in the psychology of human intelligence* (Vol. 4). Hillsdale: Lawrence Erlbaum Associates.

Sternberg, R. J. (1988b). *The triarchic mind: a new theory of intelligence*. New York: Viking.

Sternberg, R. J. (1990). *Metaphors of mind*. New York: Cambridge University Press.

Sternberg, R. J. (1993). *Sternberg triarchic abilities test*. Unpublished test.

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

Sternberg, R. J. (1997b). The triarchic theory of intelligence. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests, and issues* (pp. 92–104). New York: Guilford Press.

Sternberg, R. J. (1998). Abilities are forms of developing expertise. *Educational Researcher, 27*(3), 11–20.

Sternberg, R. J. (2003). *Wisdom, intelligence, and creativity synthesized*. New York: Cambridge University Press.

Sternberg, R. J. (2004). Culture and intelligence. *American Psychologist, 59*(5), 325–338.

Sternberg, R. J. (2007). Intelligence and culture. In S. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology* (pp. 547–568). New York: Guilford Press.

Sternberg, R. J. (2009). The rainbow and kaleidoscope projects: a new psychological approach to undergraduate admissions. *European Psychologist, 14*, 279–287.

Sternberg, R. J. (2010). *College admissions for the 21st century*. Cambridge: Harvard University Press.

Sternberg, R. J. (2015a). Multiple intelligences in the new age of thinking. In S. Goldstein, D. Princiotta, & J. Naglieri (Eds.), *Handbook of intelligence: evolutionary theory, historical perspective, and current concepts* (pp. 229–242). New York: Springer.

Sternberg, R. J. (2015b). The rainbow project and beyond: using a psychological theory of intelligence to improve the college admissions process. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world* (2nd ed., pp. 139–146). New York: Worth.

Sternberg, R. J. (2016). *What universities can be: a new model for preparing students for active concerned citizenship and ethical leadership*. Ithaca: Cornell University Press.

Sternberg, R. J. (2017). ACCEL: a new model for identifying the gifted. *Roeper Review, 39*(3), 139–152.

Sternberg, R. J., Bonney, C. R., Gabora, L., & Merrifield, M. (2012). WICS: a model for college and university admissions. *Educational Psychologist, 47*(1), 30–41.

Sternberg, R. J., & Coffin, L. A. (2010). Kaleidoscope: admitting and developing "new leaders for a changing world". *New England Journal of Higher Education, Winter, 24*, 12–13.

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

Sternberg, R. J., & Gastel, J. (1989a). Coping with novelty in human intelligence: an empirical investigation. *Intelligence, 13*, 187–197.

Sternberg, R. J., & Gastel, J. (1989b). If dancers ate their shoes: inductive reasoning with factual and counterfactual premises. *Memory and Cognition, 17*, 1–10.

Sternberg, R. J., & Grigorenko, E. L. (2001). All testing is dynamic testing. *Issues in Education, 7*(2), 137–170.

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

Sternberg, R. J., & Grigorenko, E. L. (Eds.). (2004a). *Culture and competence*. Washington, DC: American Psychological Association.

Sternberg, R. J., & Grigorenko, E. L. (2004b). Successful intelligence in the classroom. *Theory Into Practice, 43*, 274–280.

Sternberg, R. J., Grigorenko, E. L., & Kidd, K. K. (2005). Intelligence, race, and genetics. *American Psychologist, 60*, 46–59.

Sternberg, R. J., Grigorenko, E. L., Ngorosho, D., Tantufuye, E., Mbise, A., Nokes, C., Jukes, M., & Bundy, D. A. (2002). Assessing intellectual potential in rural Tanzanian school children. *Intelligence, 30*, 141–162.

Sternberg, R. J., & Hedlund, J. (2002). Practical intelligence, g, and work psychology. *Human Performance, 15*(1/2), 143–160.

Sternberg, R. J., & Kaufman, S. B. (Eds.). (2011). *Cambridge handbook of intelligence*. New York: Cambridge University Press.

Sternberg, R. J., Lipka, J., Newman, T., Wildfeuer, S., & Grigorenko, E. L. (2007). Triarchically-based instruction and assessment of sixth-grade mathematics in a Yup'ik cultural setting in Alaska. *International Journal of Giftedness and Creativity, 21*(2), 6–19.

Sternberg, R. J., Nokes, K., Geissler, P. W., Prince, R., Okatcha, F., Bundy, D. A., & Grigorenko, E. L. (2001). The relationship between academic and practical intelligence: a case study in Kenya. *Intelligence, 29*, 401–418.

Sternberg, R. J., & The Rainbow Project Collaborators. (2006). The rainbow project: enhancing the SAT through assessments of analytical, practical and creative skills. *Intelligence, 34*(4), 321–350.

Sternberg, R. J., & Suben, J. (1986). The socialization of intelligence. In M. Perlmutter (Ed.), *Perspectives on intellectual development: vol. 19, Minnesota symposia on child psychology* (pp. 201–235). Hillsdale: Lawrence Erlbaum Associates.

Visser, B. A., Ashton, M. C., & Vernon, P. A. (2006). Beyond *g*: putting multiple intelligences theory to the test. *Intelligence, 34*, 487–502.

Vygotsky, L. (1980). *Mind in society: the development of higher psychological processes*. Cambridge: Harvard University Press.

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*, 456–467.

Yang, S., & Sternberg, R. J. (1997). Taiwanese Chinese people's conceptions of intelligence. *Intelligence, 25*, 21–36.