

Full reference:

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

## **WICS: A Model for College and University Admissions**

Robert J. Sternberg  
Christina Rhee Bonney  
Liane Gabora<sup>1</sup>

Tufts University

We are grateful to the Rainbow Project Collaborators and the Kaleidoscope Project Collaborators for making this research possible. We especially want to thank Elena Grigorenko, Linda Jarvin, and Steven Stemler. Preparation of this article was supported by CASL - IES grant R305H030281, ROLE – NSF grant REC 440171, and REESE - NSF grant REC 0633952. The Rainbow Project was supported by the College Board, and the Kaleidoscope Project has been supported by Tufts University and private donations to Tufts. Correspondence concerning this article should be sent to Robert J. Sternberg, Dean of the School of Arts and Sciences, Tufts University, 3<sup>rd</sup> Floor, Ballou Hall, Medford, MA 02155, USA; robert.sternberg@tufts.edu.

### Abstract

This article outlines shortcomings of currently-used university admissions tests, and discusses ways in which they could potentially be improved, summarizing two projects designed to enhance college and university admissions. The projects were inspired by the augmented theory of successful intelligence, which postulates that intelligence comprises creative skills in generating novel ideas, analytical skills in discerning whether they are good ideas, practical skills in implementing the ideas and persuading others of their worth, and wisdom-based skills in ensuring that one uses one's knowledge and abilities in the service of a common good. In the Rainbow Project, it was possible to substantially increase prediction of first-year university academic performance and simultaneously reduce ethnic-group differences on a predictive test, relative to a standardized test used for admissions in the United States. In the Kaleidoscope Project, students admitted for expanded skills performed as well as other students, without the ethnic-group differences typically obtained in such measures. Enhanced prediction of active-citizenship and leadership activities is also demonstrated through these measures. Both projects show that it is possible to apply the theory of successful intelligence in ways that enhance the college and university admissions process. Limitations of the projects are discussed.

## **WICS: A Model for College and University Admissions**

Is it possible to create reliable, valid assessments that can be used for college and university admissions that measure not only the analytical skills measured by conventional tests, but also the creative, practical, and wisdom-based skills necessary for active citizenship and leadership throughout one's life? We set out to answer this question.

For roughly 100 years, since the work of Alfred Binet and Theodore Simon, testing to identify abilities prerequisite to academic and other forms of success has changed relatively little (see Binet & Simon, 1916). These days, if any technology other than testing were to stay about the same for 100 years, many people would be surprised. Imagine if we had only telegraphs operated by Morse code, primitive telephones, no televisions, no computers, and no serious electrical appliances. That is a world hard to imagine. Yet, it is the world in which we live in the field of ability and achievement testing.

Moreover, assessments have not stayed the same because they are optimal; there are clear signs that the tests we have been using for college and university admissions for so long are incomplete. First, they consist of questions that are somewhat contrived and artificial, and very different from the more open-ended challenges that one encounters in everyday life and that must be resolved to be successful in the real world. Second, some of these tests, such as those administered to law schools, have been shown to favor certain groups. For example, there is evidence that the test that currently dominates law school admissions decisions, the Law School Admission Test (LSAT), favors some ethnic groups over others (Kidder, 2000, 2003; Wightman, 1997, Zwick, 2002). Third, these tests do not assess as broad a base of cognitive skills as might be optimal, as we argue below.

There have been new developments seeking the understanding and assessment of a broader base of cognitive skills. Joseph Renzulli (2005), Howard Gardner (2006), and others (see Sternberg & Davidson, 2005) have proposed new models that have been used to identify abilities other than conventional intelligence. But the tests used to measure intelligence have not changed substantially. They continue to measure the same basic construct of "general ability" that Charles Spearman identified early in the twentieth century (Spearman, 1927). Our efforts have been addressed toward developing new kinds of tests to assess abilities in broader ways than has been possible in the past. This article describes the theoretical framework and some of the empirical research behind two such efforts.

The framework used here is one called the augmented theory of successful intelligence, or WICS, which is an acronym for wisdom, intelligence, creativity, synthesized (Sternberg, 1997, 1999b, 2003, 2005b; Sternberg & Grigorenko, 2004). The basic idea is that people in almost any walk of life need (a) *creativity* to generate new and exciting ideas, (b) *analytical intelligence* to evaluate whether their (and others') ideas are good ideas; and (c) *practical intelligence* to execute their ideas and to persuade others of their value. They further need (d) *wisdom* in order to ensure that their abilities are being used for some kind of common good that balances their own interests with other people's and institutional interests over the short and long terms. According to the framework,

these abilities are modifiable, in some degree, rather than fixed (Dweck, 1999; Sternberg, 1999a, 2003a; Sternberg & Grigorenko, 2007).

This framework suggests that conventional tests of abilities, dating back to Binet and Simon (1916) and Spearman (1927), are not fully adequate because they so heavily emphasize analytical (as well as memory-based) abilities to the near or total exclusion of creative and practical abilities. Such tests predict a large variety of performances (Herrnstein & Murray, 1994; Jensen, 1998; Schmidt & Hunter, 1998), but perhaps not at the highest level that can be achieved.

Our understanding of intelligence is being developed and refined in important ways (e.g., Ceci, 1996; Gardner, 1983, 2006; Guilford, 1967; Thurstone, 1938). Even within theories that postulate general intelligence, a widely accepted view is that abilities are hierarchically differentiated (e.g., Carroll, 1993; Cattell, 1971; Vernon, 1967; see essays in Sternberg & Grigorenko, 2002). So a view of broad measures of intelligence fits with many theoretical frameworks. Where the theories differ is in which abilities are measured—in what kinds of abilities are considered meritorious—and in how important the abilities are considered to be beyond general intelligence (*g*).

School assessments, such as standardized tests, often emphasize analytical and memory-based skills. For example, the SAT, used in the United States, measures, among other things, analysis of reading passages and solution of mathematics problems. The A-Levels used in the United Kingdom measure memory for knowledge learned in secondary school and basic analysis of this knowledge. These memory and analytical skills are precisely the abilities in which many students of the middle- and upper middle class excel, resulting in a fairly substantial correlation between test scores and socioeconomic class (Lemann, 1999; Sternberg, 1997). Of course, there are exceptions. But on the whole, the system of selective admissions based on tests is geared to favor these students, who have had opportunities that students of the working class may not have had. The system also is stacked against students from the middle and upper middle classes who may be nontraditional learners. As such, testing has the potential advantage of creating equity by admitting students because of their abilities and achievements, and the potential disadvantage of destroying equity by favoring, on bases other than abilities and achievements, some groups of students over others.

Success in life depends on a broader range of abilities than conventional tests measure. For example, memory and analytical abilities may be sufficient to produce A's in science courses, but they are probably not sufficient to produce outstanding research, even if they are relevant, as in deciding whether one's ideas are good ones (Lubinski, Benbow, Webb, & Bleske-Rechek, 2006). In particular, outstanding researchers must be creative in generating ideas for theories and/or experiments, analytical in discerning whether their ideas are good ones, and practical in getting their ideas funded and accepted by competitive refereed journals and funding agencies. They also require a certain degree of social skill to convince others of the worth of their ideas, and collaborate with others to develop and implement them. Success in arenas other than academic research, such as business, teaching, government, politics, and the arts, requires a similar broad array of skills that make use of creativity, analytic ability, practical ability, and wisdom.

There is evidence that creativity is important to successful leadership skills (Puccio, Murdock, & Mance, 2006). A computer model suggests that creative leadership is particularly important during the initial phase of the idea generation process, and when

the creativity level of workers is low (Leijnen & Gabora, 2010). Although early studies seemed to suggest that creativity is just one form of intelligence (Getzels & Jackson, 1962), subsequent investigations indicate that they are distinct (Wallack & Kogan, 1965), and that measures of creativity but not measures of intelligence are correlated with extracurricular activities and achievements (Kogan & Pankove, 1974; Wallach & Wing, 1969). The absence of measures of creativity from standardized tests may reflect as much as any other consideration the difficulty in easily and effectively measuring creativity (Plucker & Makel, 2010). Another key factor is the historical environment in which such tests were developed, an environment that was less dynamic and thus in which the need for creative problem solving skills was not as great. Since many facets of human existence are changing more quickly than ever before, it may be more important than ever to cultivate creative thinking skills in those who have the potential to be the leaders of tomorrow. Conventional tests thus may be a good beginning in college admissions, but, over the years, they also seem to have become the end with regard to quantified assessment of abilities.

We have been involved in two related projects exploring whether broader quantitatively-based assessments might be helpful in the university admissions process. The first of these projects is the Rainbow Project, the second, the Kaleidoscope Project. Our goal here is not to present the projects in detail, which is done elsewhere (Sternberg, 2009, in press; Sternberg & Coffin, 2010; Sternberg and the Rainbow Project Collaborators, 2006). We have also done other projects that are relevant to admissions but, for lack of space, are not described here (Grigorenko et al., 2009; Hedlund et al. 2006; Stemler et al., 2006, 2009).

How does a college or university decide whom to admit to its ranks? In this article, we propose that students should be admitted to college on the basis of their potential for future leadership and active citizenship, at whatever level of society (from the family up through the world), taking into account, among other things, their having the academic knowledge and skills necessary for success in college work (Sternberg, in press). By leadership, we mean a student's potential to make a positive, meaningful, and potentially enduring difference to the world at any level, from the family up through the community, the state, the country, or the world as a whole. Thus we are using the word 'leadership' more broadly than it is often used, to refer not just to having command over others but to having a meaningful impact directly or indirectly on society.

Many current admissions processes seek understanding of a broad range of talents, but test scores and school grades, because they are quantified, may get proportionally more weight than they deserve in a holistic admissions process seeking to select the best students for admission. Current admissions assessments such as the SAT and the ACT, based largely on theories of human abilities and competencies proposed early in the twentieth century, are somewhat narrow in the spectrum of knowledge and skills they measure, and need supplementation by assessments that draw on broader theories of human abilities and competencies.

Such broader tests would help to create systems of admissions that are more nearly equitable for the individuals applying to college and would also improve outcomes—both in terms of academic quality and diversity—for universities and society as well. In particular, students from diverse SES backgrounds may have substantial hidden talents that are relevant for success in colleges and universities. Although

admissions and instruction/assessment are often run in institutions as distinct and autonomous operations, they need to be coordinated so that instruction matches the abilities selected for in admissions. For example, if a college seeks creative students through its admissions program, then it must also assist instructors to teach in ways that encourage and value creative thinking, lest those who are admitted find themselves later in an environment incompatible with the very skills for which they were admitted.

### **What is Wrong?**

There is something lacking or missing in what we are doing in college and university admissions. We are not admitting people as we ideally should and could, and the academic disciplines are not doing an ideal job of cultivating students' skills, including their ethical ones. We wish to underscore the point: It is not that we are doing a bad job; it is rather that we can do much better. Our goal in this article is not to devalue the work of testing companies, admissions officers, or any other target group. Rather, it is to argue that our society, as a whole, has created a system of interlocking parts that do not work together as well as they could; however, it will be hard to fix because so many people do not even recognize the existence of a problem.

Until the 1960s, the main way that children were admitted to selective colleges and universities was on the basis of the social class of their parents. It was believed that one's social class would predict in great measure one's potential for future positive leadership. In the 1960s, Inslee Clark, Dean of Admissions at Yale, and others involved in college admissions, had a new vision—a vision of an elite chosen primarily on the basis of merit rather than largely on privilege. No longer would it be enough merely to come from an affluent family, or merely to have a private school education at one of the elite independent schools in the nation. Some great students might come from well-endowed families and elite private schools, but then, some might come from very poor families and public schools with few resources. So the colleges started weighing test scores and grades more heavily, and parental wealth and privilege less heavily. The result was a new generation of leaders intended to be chosen for merit rather than the luck of the draw at birth.

Given the leaders we have, ones who led Lee Iacocca (2008) to write a book asking *Where have all the leaders gone?*, we have to question whether something went wrong somewhere along the line. Actually, several things went wrong and still are wrong.

### **Problems in Producing Leaders**

**Testing and social class.** The first problem was that test scores correlate highly, although certainly not perfectly, with social class (Sacks, 2007). This correlation is not fixed, but rather, malleable. Parents of students from upper socioeconomic classes, inadvertently increase this correlation. They, unlike the less well-off social classes, can afford to spend substantial sums of money tutoring their students for the tests they will have to take, so that these students come to have an edge in the testing sweepstakes. It is understandable that well-off parents would pay for such tutoring: They are trying to do the best for their students. Who wouldn't? If only test scores well predicted future leadership, then we might be back to our hope that family legacy could well predict who will make a positive difference to the world of the future. But they are no magic bullet, for predicting future leadership or anything else. Like many other predictors, they have some predictive value, but that value is not very high.

This problem is a specific example of a more general principle sometimes known as a Matthew effect. The Matthew effect derives from a statement in the Bible:

*For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath." ([Matthew XXV:29](#)).*

The statement is another way of saying that the rich get richer, and the poor get poorer. The late Robert Merton (1968), a sociologist at Columbia University, applied this principle to scientists: to those at well-known schools who already had a good reputation came more resources; to those at poorly-known schools who had little reputation, came little, resulting in the further derailment of their careers. But the Matthew principle applies as well to college admissions. If your parents can send you to strong schools and pay for tutelage, they can help bring you to a position where you are more competitive for college admissions; if your parents cannot pay, it becomes increasingly difficult to gain access on the basis of traditional measures of merit. Good admissions officers, however, take these factors into account and consider access to resources when evaluating applications.

Why should a correlation of test scores with socioeconomic status even be a problem? After all, if students from higher socioeconomic strata are able to perform better in college then why not simply let them in, given that, in that case, the college-admissions tests would be accurately predicting their performance?

There is no simple answer to this question and the answers one provides are a matter of policy, not psychology. However, we believe there are three reasons why the correlation is problematical. In discussing these reasons, we will draw on an imaginary society. This society educates well people with blue eyes but poorly educates people with brown eyes, having decided that eye color is a good measure of the educational and perhaps other resources to which a person is entitled. (In our society, of course, historically we have used gender and skin color, which are not clearly better measures!) When the children reach high school, society gives them a test. The blue-eyed children outscore the brown-eyed children by a large margin and then, because college admissions are largely determined by test scores, the blue-eyed children dominate the numbers in college.

The first reason is that if one makes decisions largely on test scores, as partial proxies for socioeconomic status (which itself may be a proxy for other things), then one reduces the chances of creating a more equitable society. There will be those who will say that society has been fair, because it gave preference to those who deserved it. Unsurprisingly, many, but not all, who think in that way will have blue eyes. Many of the brown-eyed adults may say that their children were never given a chance, but the brown-eyed adults will have less money and power in society, and will not be well heard. Moreover, some of those who do hear them will view their complaints as sour grapes. Without some kind of admissions program that looks for strengths in the brown-eyed children, the society will be stuck with a hereditary hierarchy that seems merit-based—at least to those with blue eyes, and others who are perfectly rational on many matters not pertaining to the continued power and well-being of themselves and their children. [portraying perceptions as established facts? Reference needed?]

The second reason is that the brown-eyed children may actually be better in some skills than the blue-eyed children. This superiority will probably not show up in assessments created by typical blue-eyed adults, who have come to view as valuable

those skills in which they and their children excel. But without efforts to devise measures that may do justice to the full range of skills of *all* children, one may in effect create tests skewed to favor those with blue eyes. Moreover, the teachers, also primarily of blue eyes or at least trained by those with blue eyes, will value the same narrower range of skills. So they will construct assessments that match the college admissions test, in the same way that the college admissions test may well originally have been created to value what the blue-eyed teachers, or those teachers trained by those with blue eyes, value. So the tests will seem fair because the system is created to value what it is that blue-eyed people, on average, do well, but not what brown-eyed people, on average, do well.

The third reason, implied above, is that the criterion may be as biased, in a sense, as the predictor. When people are hired for upper level jobs, often what employers look for, among other things, are what college the student went to (e.g., Harvard U. vs. Stinky U.) and what kinds of grades the student received at the college. But schools and grades are a highly imperfect predictor of job performance. The recent debacle on Wall Street is an example. It was created largely by individuals with excellent grades at terrific business schools. It was their ascent into positions of great power that enabled them to bring down not only Wall Street, but much of the economic system of the United States and the world. It is a tribute to a system so stacked in favor of the “blue eyes” that, when some of them were fired for extraordinarily incompetent performance, they received severance packages—and many still there are receiving bonuses—in excess of what many brown-eyed people will make over the course of their lives. Heads they win; tails, others lose.

It is important to note in this consideration that high school grades, like SATs, have correlations with socioeconomic status that may be as great as or even greater than those of the standardized tests (Zwick, 2002; Zwick & Green, 2007). So ditching the tests in favor exclusively of school grades might solve some problems, but not that of the correlation of academic performances with socioeconomic status. The correlation of socioeconomic status with high school performance is not terribly surprising. Educated parents can help their high school student children in ways that less educated parents cannot.

So let's consider other problems with the present system.

**Narrowness of assessments.** The second problem was that the tests were narrow in their conception. In the early 20<sup>th</sup> century, when the tests were first conceived, perhaps the abstract, academic skills they measured served as a somewhat reasonable basis for distinguishing more able white males of privilege from less able ones. But now, in the early 21st century, the skill set one needs for success in college and life has substantially broadened. Academic knowledge alone would not get one through. The world simply changes too quickly. For this reason, many colleges use holistic admissions practices that take into account credentials beyond just academic knowledge.

It is worth emphasizing that the system we have now was developed to create equity, not to destroy it. In earlier times, the main way one got ahead was on the strength of one's family connections. Tests were designed to create an objective measure of potential that would move beyond family ties. The procedures may have made sense in the early twentieth century, when the only people taking the tests were the children of privilege, so one could distinguish better among them who had more potential. But as the range of students taking the tests increased, and their backgrounds became more diverse, what had worked in the early twentieth century simply no longer worked as well. One

could no longer assume that almost all the test-takers had similar upbringings, backgrounds, and opportunities for schooling. Rather, their opportunities were diverse and so new tests needed to reflect that. But they didn't.

**Secondary school as test preparation.** The third problem is that the tests transformed secondary schooling, and not necessarily for the better. Students have come to spend more and more time preparing for tests, and less and less time learning lessons that may be meaningful to them in later life. Education has come, in some measure, to be replaced by gamesmanship. Music, art, physical education, and even, in some schools, social studies and science, have gone by the wayside if they are not formally tested by the states. [Reviewer 3 says that references are needed here. Bob?]

If the view of this article is correct—that students from diverse SES backgrounds may have substantial hidden talents that are relevant for success in colleges and universities—how can one show it? The documentation of what some people believe intuitively has been something of a stumbling block for merit-based alternatives or supplements to affirmative action. We have attempted to derive such a system, based on the theory of successful intelligence. Such a theory can show hidden talents not only in students from lower SES backgrounds, but in those from upper SES backgrounds who do not learn and think in traditional ways.

Our view that we might have success in this realm dates back at least to a study in which Sternberg, Ferrari, Clinkenbeard, and Grigorenko (1996) gave a test that they had devised to over 300 high school students across the United States in order to select students with different patterns of strengths on the basis of analytical, creative, and practical abilities. In particular, they divided them into high-analytical, high-creative, high-practical, high-analytical/creative/practical, and low-analytical/creative/practical groups. The identification was prior to their being placed into sections to take a college-level summer psychology course. When they divided the students into such groups, they noticed something that, at the time, was unexpected. Students in the high-analytical group, who excelled in the abilities measured by conventional tests, were for the most part white and middle-class. Students in the high creative and high practical groups were ethnically diverse and diverse in their geographic origins.

They found that when students were taught in a way that matched their patterns of abilities, at least some of the time, they excelled. In other words, the creatively- and practically-gifted students did excel, so long as, at least some of the time, the way they were taught matched the way they learned. Good teachers use a variety of teaching methods to reach diverse learning styles of their students, so any student taught in a way that is responsive to his or her pattern of abilities can excel. But traditional assessments for college readiness tend to emphasize the memory-analytical mode of learning, to the exclusion of others.

After this study, Sternberg and his colleagues went on to show that teaching to all styles of learning does indeed improve achievement relative to teaching that emphasizes just traditional memory-analytical patterns of learning and thinking (Sternberg, Torff, & Grigorenko, 1998). But the seeds of a further question were planted in us: Is it possible that many students who are not now being identified as having impressive credentials for college or graduate work might in fact be so identified if they were assessed in a way that looked at creative and practical, as well as analytical forms of skills? The Rainbow

Project sought to answer this question. The Kaleidoscope Project added wisdom as a quality to be considered.

### **Projects to Broaden the Spectrum of Admissions**

#### **The Rainbow Project**

When universities make decisions about selective admissions, the main quantitative information they have available to them typically is grade-point average in high school or its equivalent and scores on standardized tests (Lemann, 1999). Is it possible to create assessments that are psychometrically sound and provide incremental validity over existing measures, without destroying the cultural and ethnic diversity that makes a university environment a place in which students can interact with and learn from others who are different from themselves? Can one create assessments recognizing that people's gifts differ and that many of the variety of gifts they possess are potentially relevant to university and life success (Sternberg & Davidson, 2005)? In addition, can one do so in a way that is not a mere proxy for socioeconomic class (Golden, 2006; Kabaservice, 2004; Karabel, 2006; Lemann, 1999; McDonough, 1997) or for IQ (Frey & Detterman, 2004)?

The Rainbow Project (for details, see Sternberg & the Rainbow Project Collaborators, 2006; see also Sternberg, 2005a, 2006; Sternberg & the Rainbow Project Collaborators, 2005; Sternberg, the Rainbow Project Collaborators, & the University of Michigan Business School Collaborators, 2004) was a first project designed to enhance university admissions procedures at the undergraduate level. The Rainbow measures were intended to supplement the SAT, but they can supplement any conventional standardized test of abilities or achievement. In the theory of successful intelligence, abilities and achievement are viewed as being on a continuum—abilities are largely achieved (Sternberg, 1998a, 1999a)—so it is not clear that it matters greatly exactly what test is used, given that most of the tests that are used are highly *g*-loaded.

The SAT is a comprehensive examination currently measuring verbal comprehension and mathematical thinking skills, with a writing component recently added. A wide variety of studies have shown the utility of the SAT and similar tests as predictors of university and job success, with success in college typically measured by GPA (grade-point average) (Hezlett et al., 2001; Kobrin et al., 2002; Schmidt & Hunter, 1998). Taken together, these data suggest reasonable predictive validity for the SAT in predicting undergraduate performance. Indeed, traditional intelligence or aptitude tests have been shown to predict performance across a wide variety of settings. But as is always the case for a single test or type of test, there is room for improvement. The theory of successful intelligence provides one basis for improving prediction and possibly for establishing greater equity and diversity, which is a goal of most higher-educational institutions (Bowen, Kurzweil, & Tobin, 2006). It suggests that broadening the range of skills tested to go beyond analytic skills, to include practical and creative skills as well, might significantly enhance the prediction of undergraduate performance beyond current levels. Thus, the theory does not suggest *replacing*, but rather, *augmenting* the SAT and similar tests such as the ACT or the A-Levels in the undergraduate-admissions process. A collaborative team of investigators sought to study how successful such an augmentation could be. Even if we did not use the SAT, ACT, or A-Levels, in particular, we still would need some kind of assessment of the memory and analytical abilities the tests assess.

#### ***Methodological Considerations***

In the Rainbow Project, data were collected at 15 schools across the United States, including 8 four-year undergraduate institutions, 5 community colleges, and 2 high schools.

The participants were 1,013 students predominantly in their first year as undergraduates or their final year of high school. In this report, analyses only for undergraduate students are discussed because they were the only ones for whom the authors had data available regarding undergraduate academic performance. The final number of participants included in these analyses was 793.

Baseline measures of standardized test scores and high-school grade-point average were collected to evaluate the predictive validity of current tools used for undergraduate admission criteria, and to provide a contrast for the current measures. Students' scores on standardized university entrance exams were obtained from the College Board.

The measure of analytical skills was provided by the SAT plus multiple-choice analytical items we added measuring inference of meanings of words from context, number series completions, and figural matrix completions.

Creative skills were measured by multiple-choice items and by performance-based items. The multiple-choice items were of three kinds. In one, students are presented with verbal analogies preceded by counterfactual premises (e.g., money falls off trees). They have to solve the analogies as though the counterfactual premises were true. In a second, students are presented with rules for novel number operations, for example, "flix," which involves numerical manipulations that differ as a function of whether the first of two operands is greater than, equal to, or less than the second. Participants have to use the novel number operations to solve presented math problems. In a third, participants are first presented with a figural series that involves one or more transformations; they then have to apply the rule of the series to a new figure with a different appearance, and complete the new series. These measures are not typical of assessments of creativity and were included for relative quickness of participants' responses and relative ease of scoring.

Creative skills also were measured using open-ended measures. One measure required writing two short stories with a selection from among unusual titles, such as "The Octopus's Sneakers," one required orally telling two stories based upon choices of picture collages, and the third required captioning cartoons from among various options. Open-ended performance-based answers were rated by trained raters for novelty, quality, and task-appropriateness. Multiple judges were used for each task and satisfactory reliability was achieved (Sternberg & the Rainbow Project Collaborators, 2006).

Multiple-choice measures of practical skills were of three kinds. In one, students are presented with a set of everyday problems in the life of an adolescent and have to select the option that best solves each problem. In another, students are presented with scenarios requiring the use of math in everyday life (e.g., buying tickets for a ballgame), and have to solve math problems based on the scenarios. In a third, students are presented with a map of an area (e.g., an entertainment park) and have to answer questions about navigating effectively through the area depicted by the map.

Practical skills also were assessed using three situational-judgment inventories: the Everyday Situational Judgment Inventory (Movies), the Common Sense Questionnaire, and the College Life Questionnaire, each of which tap into different types of tacit knowledge. The general format of tacit-knowledge inventories has been described

in Sternberg et al. (2000), so only the content of the inventories used in this study will be described here. The movies presented everyday situations that confront undergraduate students, such as asking for a letter of recommendation from a professor who shows, through nonverbal cues, that he does not recognize you very well. One then has to rate various options for how well they would work in response to each situation. The Common Sense Questionnaire provided everyday business problems, such as being assigned to work with a coworker whom one cannot stand, and the College Life Questionnaire provided everyday university situations for which a solution was required.

Unlike the creativity performance tasks, in the practical performance tasks the participants were not given a choice of situations to rate. For each task, participants were told that there was no “right” answer, and that the options described in each situation represented variations on how different people approach different situations.

Consider examples of the kinds of items one might find on the Rainbow Assessment. An example of a creative item might be to write a story using the title “3516” or “It’s Moving Backward.” Another example might show a collage of pictures in which people are engaged in a wide variety of activities helping other people. One would then orally tell a story that takes off from the collage. An example of a practical item might show a movie in which a student has just received a poor grade on a test. His roommate had a health crisis the night before, and he had been up all night helping his roommate. His professor hands him back the test paper, with a disappointed look on her face, and suggests to the student that he study harder next time. The movie then stops. The student then has to describe how he would handle the situation. Or the student might receive a written problem describing a conflict with another individual with whom she is working on a group project. The project is getting mired down in the interpersonal conflict. The student has to indicate how she would resolve the situation to get the project done. All materials were administered in either of two formats. A total of 325 of the university students took the test in paper-and-pencil format, whereas a total of 468 students took the test on the computer via the World Wide Web.

No strict time limits were set for completing the tests, although the instructors were given rough guidelines of about 70 minutes per session. The time taken to complete the battery of tests ranged from two to four hours. As a result of the lengthy nature of the complete battery of assessments, participants were administered parts of the battery using an intentional incomplete overlapping design. The participants were randomly assigned to the test sections they were to complete. Details of the use of the procedure are in Sternberg and the Rainbow Project Collaborators (2006).

Creativity in this (and the subsequent Kaleidoscope) project was assessed on the basis of the novelty and quality of responses. Practicality was assessed on the basis of the feasibility of the products with respect to human and material resources, and the persuasiveness of the product.

### ***Findings***

The analysis described below is a conservative one that does not correct for differences in the selectivity of the institutions at which the study took place. In a study across so many undergraduate institutions differing in selectivity, validity coefficients will seem to be lower than is typical, because an A at a less selective institution counts the same as an A at a more selective institution. When the authors corrected for institutional selectivity, the results described below became stronger. But correcting for

selectivity has its own problems (e.g., on what basis does one evaluate selectivity?), and so uncorrected data are used in this report. We also did not control for university major: Different universities may have different majors, and the exact course offerings, grading, and populations of students entering different majors may vary from one university to another, rendering control difficult.

When examining undergraduate students alone, the sample showed a slightly higher mean level of SAT than that found in undergraduate institutions across the United States. The standard deviation was above the normal 100-point standard deviation, meaning we did not suffer from restriction of range. Our means, although slightly higher than typical, are within the range of average undergraduate students. Another potential concern is pooling data from different institutions. We pooled data because in some institutions we simply did not have large enough numbers of cases for the data to be meaningful.

Some scholars believe that there is only one set of skills that is highly relevant to school performance, what is sometimes called “general ability,” or *g* (e.g., Jensen, 1998). These scholars believe that tests may appear to measure different skills, but when statistically analyzed, show themselves just to be measuring the single general ability. Does the test actually measure distinct analytical, creative, and practical skill groupings? Factor analysis addresses this question. Three meaningful factors were extracted from the data: practical performance tests, creative performance tests, and multiple-choice tests (including analytical, creative, and practical). In other words, multiple-choice tests, regardless of what they were supposed to measure, clustered together (see also Sternberg, Castejón, Prieto, Hautamäki, & Grigorenko, 2001, for similar findings). Thus, method variance proved to be very important. The results show the importance of measuring skills using multiple formats, precisely because method is so important in determining factorial structure. The results also show the limitations of exploratory factor analysis in analyzing such data, and of dependence on multiple-choice items outside the analytical domain. In the ideal, one wishes to ensure that one controls for method of testing in designing aptitude and other test batteries.

Undergraduate admissions offices are not interested, exactly, in whether these tests predict undergraduate academic success. Rather, they are interested in the extent to which these tests predict school success *beyond* those measures currently in use, such as the SAT and high school grade-point-average (GPA). In order to test the incremental validity provided by Rainbow measures above and beyond the SAT in predicting GPA, a series of statistical analyses (called hierarchical regressions) was conducted that included the items analyzed above in the analytical, creative, and practical assessments.

If one looks at the simple correlations, the SAT-V, SAT-M, high school GPA, and the Rainbow measures all predict first-year undergraduate GPA. But how do the Rainbow measures fare on incremental validity? In one set of analyses, the SAT-V, SAT-M, and high school GPA were included in the first step of the prediction equation because these are the standard measures used today to predict undergraduate performance. Only high school GPA contributed uniquely to prediction of undergraduate GPA. Inclusion of the Rainbow measures roughly doubled prediction (percentage of variance accounted for in the criterion) versus the SAT alone. These results suggest that the Rainbow tests add considerably to the prediction gotten by SATs alone. They also suggest the power of high

school GPA in prediction, particularly, because it is an atheoretical composite that includes within it many variables, including motivation and conscientiousness.

Studying group differences requires careful attention to methodology and sometimes has led to erroneous conclusions (Hunt & Carlson, 2007). Although one important goal of the Rainbow Project was to predict success in the undergraduate years, another important goal involved developing measures that reduce ethnic group differences in mean levels. There has been a lively debate as to why there are socially-defined racial group differences, and as to whether scores for members of underrepresented minority groups are over- or under-predicted by SATs and related tests (see, e.g., Bowen & Bok, 2000; Camara & Schmidt, 1999; Rowe, 2005; Rushton & Jensen, 2005; Sternberg, Grigorenko, & Kidd, 2005; Turkheimer et al., 2003). There are a number of ways one can test for group differences in these measures, each of which involves a test of the size of the effect of ethnic group. Two different measures were chosen:  $\omega^2$  -- omega squared and Cohen's D.

There were two general findings. First, in terms of overall differences, the Rainbow tests appeared to reduce ethnic-group differences relative to traditional assessments of abilities like the SAT. Second, in terms of specific differences, it appears that the Latino students benefited the most from the reduction of group differences. The black students, too, seemed to show a reduction in difference from the white mean for most of the Rainbow tests, although a substantial difference appeared to be maintained with the practical performance measures.

Although the group differences are not perfectly reduced, these findings suggest that measures can be designed that reduce ethnic and racial group differences on standardized tests, particularly for historically disadvantaged groups like black and Latino students. These findings have important implications for reducing adverse impact in undergraduate admissions.

The SAT is based on a conventional psychometric notion of cognitive skills. Using this notion, it has had substantial success in predicting undergraduate academic performance. The Rainbow measures alone roughly doubled the predictive power of Undergraduate GPA when compared to the SAT alone. Additionally, the Rainbow measures predict substantially beyond the contributions of the SAT and High School GPA. These findings, combined with encouraging results regarding the reduction of between-ethnicity differences, make a compelling case for furthering the study of the measurement of analytic, creative, and practical skills for predicting success in the university.

One important goal for projects such as Rainbow, which are based on the successful intelligence framework, is the creation of standardized assessments that reduce the different outcomes between different groups as much as possible to maintain test validity. The measures described here suggest results toward this end. Although the group differences in the tests were not reduced to zero, the tests did substantially attenuate group differences relative to other measures such as the SAT. This finding could be an important step toward ultimately ensuring fair and equal treatment for members of diverse groups in the academic domain.

The principles behind the Rainbow Project apply at other levels of admissions as well. For example, Hedlund, Wilt, Nebel, Ashford, and Sternberg (2006) have shown that the same principles can be applied in admissions to business schools, also with the result

of increasing prediction and decreasing ethnic- (as well as gender-) group differences. Stemler, Grigorenko, Jarvin, and Sternberg (2006) have found that including creative and practical items in augmented psychology and statistics Advanced Placement Examinations can reduce ethnic-group differences on the tests. And the same principles are being employed in a test for identification of gifted students in elementary school (Chart, Grigorenko, & Sternberg, 2008).

It is one thing to have a successful research project, and another actually to implement the procedures in a high-stakes situation. We have had the opportunity to do so. The results of a second project, Project Kaleidoscope, are reviewed here.

### **The Kaleidoscope Project**

Tufts University in Medford, Massachusetts, USA, has strongly emphasized the role of active citizenship in education. It has put into practice some of the ideas from the Rainbow Project. In collaboration with Dean of Admissions Lee Coffin, we instituted Project Kaleidoscope, which represents an implementation of the ideas of Rainbow, but goes beyond that project to include in its assessment the construct of wisdom (for more details, see Sternberg, 2007b, 2007c, 2009, in press; Sternberg & Coffin, in press).

Over 15,000 students apply for undergraduate admission at Tufts each year, either to the School of Arts and Sciences, or to the School of Engineering. Questions were added to the admissions application for all undergraduates admissions, which were designed to assess wisdom, analytical and practical intelligence, and creativity synthesized (WICS), an extension of the theory of successful intelligence (Sternberg, 2003b). The program was continued for 2007-2008, 2008-2009, and 2009-2010, but the data reported here are primarily for the first year and secondarily for the second year, for which we have the most nearly complete data.

The WICS model augments the original theory of successful intelligence (Sternberg, 1997) such that the balance theory of wisdom (Sternberg, 1998b) unifies the other intelligence components into a model of successful leadership. The main idea behind the WICS model is that wisdom is the application of intelligence, creativity, and knowledge for the common good, by balancing intrapersonal, interpersonal, and extrapersonal interests, over the long and short terms, through the infusion of positive ethical values.

The questions on the Tufts admission application are optional. Whereas the Rainbow Project was done as a separate high-stakes test administered with a proctor, the Kaleidoscope Project has been done as a section of the Tufts-specific supplement to the Common Application. It is simply not practical to administer a separate high-stakes test such as the Rainbow assessment for admission to one university. Moreover, the advantage of Kaleidoscope is that it removed the pressure of a high-stakes testing situation in which students must answer complex questions in very short amounts of time under incredible pressure. Instead, since it was part of their college application, students were able to take as much time as they needed to prepare their response and make revisions.

Students were encouraged to answer just a single question out of roughly eight to ten possible choices so as not overly to burden them. Tufts University competes for applications with many other universities, and if our application was substantially more burdensome than those of our competitor schools, it would put us at a real-world disadvantage in attracting applicants. In the augmented theory of successful intelligence,

successful intelligent individuals capitalize on strengths and compensate for or correct weaknesses. Our format gave students a chance to capitalize on a strength.

The items, for the most part, change each year. As examples of items, a creative question asked students to write stories with titles such as “The End of MTV” or “Confessions of a Middle-School Bully.” Another creative question asked students what the world would be like if some historical event had come out differently, for example, if Rosa Parks had given up her seat on the bus. Yet another creative question, a nonverbal one, gave students an opportunity to design a new product or an advertisement for a new product. A practical question queried how students had persuaded friends of an unpopular idea they held. A wisdom question asked students how a passion they had could be applied toward a common good.

Creativity and practicality were assessed in the same way as in the Rainbow Project. But there were two additional assessments. Analytical quality was assessed by the organization, logic, and balance of the essay. Wisdom was assessed by the extent to which the response represented the use of abilities and knowledge for a common good by balancing one’s own, others’, and institutional interests over the long and short terms through the infusion of positive ethical values.

Note that the goal is not to replace SAT and other traditional admissions measurements like grade point averages and class rank with some new test. Rather, it is to re-conceptualize applicants in terms of academic/analytical, creative, practical, and wisdom-based abilities, using the essays as one but not the only source of information. For example, highly creative work submitted in a portfolio also could be entered into the creativity rating, or evidence of creativity through winning of prizes or awards. The essays were major sources of information, but since these were optional, admissions officers were trained to also look for WICS characteristics in other areas of each application they reviewed, and provide a Kaleidoscope rating whenever possible.

We now have some results of our first and second years of implementation, and they are very promising (Sternberg, Bonney, & Gabora, in press). In the first year, applicants were evaluated for creative, practical, and wisdom-based skills; in subsequent years, the analytical dimension was also rated by admissions officers.

Among the applicants who were evaluated as being academically qualified for admission, approximately half completed an optional essay the first year, and about two-thirds in subsequent years. Doing these essays had no meaningful effect on chances of admissions. However, *quality* of essays or other evidence of creative, practical, or wisdom-based abilities did have an effect. For those 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 getting an A. Because of the large number of essays (roughly 8,000 in the first year and over 10,000 in subsequent years), only one rater rated applicants except for a sample to ensure that inter-rater reliability was sufficient. The Kendall Tau-b rank correlation was used to test inter-rater reliability since the Kaleidoscope ratings are ordinal measures. Across the creative, practical, and wisdom-based dimensions evaluated the first year, the tau-b statistic was .62, .58, and .56, respectively, between the pairs of raters. Essays were not a substitute for good academic credentials. Rather, they were most helpful when academic credentials were neither extremely strong nor extremely weak. In the latter case, applicants were rejected, regardless of their essays.

Many measures do not look like conventional standardized tests, but have statistical properties that mimic them. We were therefore interested in convergent-discriminant validation of our measures. The correlation of our measures with a rated academic composite that included SAT scores and high school GPA were modest but significant for creative thinking, practical thinking, and wise thinking. The correlations with a rating of quality of extracurricular participation and leadership were moderate for creative, practical, and wise thinking. Thus, the pattern of convergent-discriminant validation was what we had hoped for.

The average academic quality of applicants in Arts & Sciences rose in slightly 2006-7, the first year of the pilot, in terms of both SAT and high school grade-point average. Indeed, SATs and high school GPAs have risen in each year since Kaleidoscope was first administered. In addition, there were notably fewer students in what before had been the bottom third of the pool in terms of academic quality. Many of those students, seeing the new application, seem to have decided not to bother to apply. Many more strong applicants applied.

Thus, adopting these new methods does not result in less qualified applicants applying to the institution and being admitted. Rather, the applicants who are admitted are *more* qualified, but in a broader way. Perhaps most rewarding were the positive comments from large numbers of applicants that they felt our application gave them a chance to show themselves for who they are. Of course, many factors are involved in admissions decisions, and Kaleidoscope ratings were only one small part of the overall picture.

We did not get meaningful differences across ethnic groups, a result that surprised us, given that the earlier Rainbow Project reduced but did not eliminate differences. And after a number of years in which applications by underrepresented minorities were relatively flat in terms of numbers, the first year they went up substantially. In the end, the number of applications from African Americans has increased roughly 30% and Hispanic-American applications increased by roughly 15% since the implementation of Kaleidoscope at Tufts. The number of African American students *admitted* (not just those that apply for admission) has increased roughly 25%. So our results, like those of the Rainbow Project, showed that it is possible to increase academic quality and diversity simultaneously, and to do so in for an entire undergraduate class at a major university, not just for small samples of students at some scattered schools. Most importantly, we sent a message to students, parents, high school guidance counselors, and others, that we believe that there is a more to a person than the narrow spectrum of skills assessed by standardized tests, and that these broader skills can be assessed in a quantifiable way. A possible alternative interpretation for this result, however, is that since the admissions officers were not blind to information about race, there may have been increased awareness of affirmative action and the potential benefits of diversity. Further studies are needed to distinguish between these alternative explanations.

We found that, holding high school GPA and SATs constant, students who received Kaleidoscope ratings performed significantly better in their freshman year than did those who did not receive such ratings. We also found that the levels of ratings predicted measures of extracurricular participation and satisfaction in leadership, active-citizenship, and other forms of engagement outside the classroom.

### **Limitations**

There are many limitations of these studies that circumscribe the conclusions that can be drawn from them. A first limitation is that socioeconomic class is confounded with ethnicity. So ethnicity differences may be attributable, in unknown measure, to socioeconomic class differences. The differences are unlikely to be solely a function of socioeconomic class in that, where we obtained differences, others have obtained similar patterns of differences (see, e.g., Loehlin, Lindzey, & Spuhler, 1975). For example, Asian-Americans did better on quantitative analytical measures and worse on creative measures than did White Americans, but in a result with Chinese and American college students in comparably selective universities, we obtained the same result regarding creativity, regardless of whether we used Chinese or American university professors as raters (Niu & Sternberg, 2001; see also review in Lynn, 2006). Moreover, Asian Americans are generally not at as high a socioeconomic level than are whites but performed better on the quantitative analytical tests, such as the math SAT, here and in other studies (Lynn, 2006). The reason we did not control for socioeconomic class is that we were unable to obtain the data that would have enabled us to do so.

A second limitation is that there were problematical methodological issues in both the Rainbow and Kaleidoscope Projects. In Rainbow, we used an incomplete design, meaning that not all students took all tests. This made the statistical analysis complex to the point where we would not recommend the use of this design by others. In Kaleidoscope, unlike in Rainbow, assessments were done without proctoring. Thus, we cannot be certain of the conditions under which the assessments were taken, or even that it was the applicant who took the assessment. The nature of the assessments, though, makes it questionable whether parents or others who might take the assessment would do better than the applicants (for example, to cite one of the essays, many parents know far less about MTV than do their children). Moreover, an advantage of doing the assessment at home is that students have more time to think carefully and deeply than they do in a timed proctored test; often it is hard to think creatively, practically, or wisely without having sufficient time to do so. These methodological issues result from the fact that this was a “real-world” study rather than a strictly scientific one with matched controls. While this makes it much more difficult to make generalizable claims stemming from our project, it should not detract from the importance of the project’s implications for college admissions.

A third limitation is that the new assessments require more time, resources, and money to score the assessments. We had to hire raters and train them. Although reliability was good, it could only be achieved with training. Schools would therefore have to decide that the additional information was worth the cost. In the Rainbow Project, we got substantially better prediction than SAT alone (double) or SAT plus high school GPA (roughly 50% increase) and decrease ethnic-group differences. In the Kaleidoscope Project, we did not see academic differences between groups as function of levels of Kaleidoscope scores, but rather, as a result of being rated for Kaleidoscope.

A fourth limitation is that our follow-up data at this time are limited. For Rainbow, we had only first-year university grades. For Kaleidoscope, we have less than four full years of performance. For this project, we hope to follow up by measuring progress broadly—including nonacademic measures—during the four years the students are at the university.

A fifth limitation, in Kaleidoscope, is selection bias. Students who completed the essays were not a random sample of applicants: They chose to do extra work. However, because admission probabilities were not related to the fact of completing the essays, only to quality of essays for those who did complete them, the bias may not have been an important factor in the results. Note that it is useful to keep in mind here that Kaleidoscope is not meant to be a controlled laboratory experiment with obvious generalizability. Kaleidoscope was studied while it was being used in admissions so strict experimental controls were not possible.

Yet another criticism one could level at this approach is the following. Student performance on the SAT is correlated with parental SES, partly due to the fact that there is much preparation for taking the SAT, and preparation is correlated with SES. It could be argued that the Rainbow / Kaleidoscope measures would similarly be susceptible to effects of such preparation. However, there exist formal courses for preparing for the SAT. There are no formal courses for preparing for Rainbow / Kaleidoscope measures, and it is not as straightforward how creativity, leadership and practical skills could be improved through a preparatory test.

### **Conclusions**

There is perhaps no element of the educational system that does such a poor job of cultivating the leaders of tomorrow as the admissions process currently used to determine who gets into colleges and universities. We have examined some of the shortcomings of this process, and discussed projects designed to supplement and/or modify this process based on the theory of successful intelligence. The theory of successful intelligence appears to provide a strong theoretical basis for augmented assessment of the skills needed for undergraduate success. There is evidence to indicate that it has good incremental predictive power, and serves to increase equity. As teaching improves and university teachers emphasize more the creative, practical, and wisdom-based skills needed for success in school and life, the predictive power of the test may increase.

Cosmetic changes in testing over the last century have made relatively little difference to the construct validity of assessment procedures. The augmented theory of successful intelligence could provide a new opportunity to increase construct validity at the same time that it reduces differences in test performance between groups. It may indeed be possible to accomplish the goals of affirmative action through tests such as the Rainbow and Kaleidoscope assessments, either as supplements to traditional affirmative-action programs or as substitutes for them.

Other modern theories of intelligence, such as those mentioned earlier in the article (e.g., Ceci, 1996; Gardner, 1983), may also serve to improve prediction and increase diversity. Moreover, other approaches to supplementing the SAT, and the Rainbow tests, may be called for. For example, Oswald, Schmitt, Kim, Ramsay, and Gillespie (2004) (see also Schmitt et al., in press) have found biodata and situational-judgment tests (the latter of which we also used) to provide incremental validity to the SAT. Sedlacek (2004) has developed non-cognitive measures that appear to have had success in enhancing the university-admissions process.

The theory and principles of assessment described in this article can be extended beyond the United States (Sternberg, 2004, 2007a). We have used assessments based on the theory of successful intelligence on five continents, and found that the general

principles seem to hold, although the content used to assess abilities need to differ from one locale to another. At present, we are engaging in a collaboration with psychologists in Germany to determine whether the instruments we have used in the United States might, in suitable form, be useful there as well.

There is no question but that the methods used in the Rainbow Project, the Kaleidoscope Project, and related projects are at early stages of development. They do not have more than 100 years of experience behind them, as do traditional methods. What the results suggest is that an argument is to be made for broader assessments—that broader assessments are not synonymous with fuzzy-headed assessments. Such assessments can improve prediction and increase diversity, rather than trading off the one for the other. Broader assessments do not replace conventional ones: They supplement them. Our results show an important role for traditional analytical abilities in university success. But these are not the only abilities that matter, and should not be the only abilities we measure.

### References

- Binet, A., & Simon, T. (1916). *The development of intelligence in children*. Baltimore: Williams & Wilkins. (Originally published in 1905).
- Bowen, W. G., & Bok, D. (2000). *The shape of the river: long-term consequences of considering race in college and university admissions*. Princeton, NJ: Princeton University Press.
- Bowen, W. G., Kurzweil, M. A., & Tobin, E. M. (2006). *Equity and excellence in American higher education*. Charlottesville, VA: University of Virginia Press.
- Camara, W. J., & Schmidt, A. E. (1999). *Group differences in standardized testing and social stratification* (College Board Research Rep. No. 99-5). New York, NY: The College Board. Retrieved 12/21/2006 from <http://www.collegeboard.com/research/home/>.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. NY: World Book Co.
- Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton Mifflin.
- Ceci, S. J. (1996). *On intelligence: A bioecological treatise on intellectual development*. Cambridge, MA: Harvard University Press.
- Chart, H., Grigorenko, E. L., & Sternberg, R. J. (2008). Identification: The Aurora Battery. In J. A. Plucker & C. M. Callahan (Eds.), *Critical issues and practices in gifted education* (pp. 281-301). Waco, TX: Prufrock.
- Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia: Psychology 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*, 373-378.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic.
- Gardner, H. (2006). *Multiple intelligences: New horizons*. New York: Perseus.
- Getzels, J. W. & Jackson, P. W. (1962). *Creativity and intelligence: Explorations with gifted students*. New York: Wiley.

- Golden, D. (2006). *The price of admission*. New York: Crown.
- Grigorenko, E. L., Jarvin, L., Diffley, R., Goodyear, J., Shanahan, E. J., & Sternberg, R. J. (2009). Are SSATs and GPA enough? A theory-based approach to predicting academic success in high school. *Journal of Educational Psychology, 101*, 964-981.
- Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- 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.
- Herrnstein, R. J., & Murray, C. (1994). *The bell curve*. New York: Free Press.
- Hezlett, S., Kuncel, N., Vey, A., Ones, D., Campbell, J., & Camara, W. J. (2001). *The effectiveness of the SAT in predicting success early and late in college: A comprehensive meta-analysis*. Paper presented at the annual meeting of the National Council of Measurement in Education, Seattle, WA.
- Hunt, E., & Carlson, J. (2007). Considerations relating to the study of group differences in intelligence. *Perspectives on Psychological Science, 2*, 194-213.
- Iacocca, L. (2008). *Where have all the leaders gone?* New York: Scribner.
- Jensen, A. R. (1982). The chronometry of intelligence. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 255-310). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Jensen, A. R. (1998). *The g factor*. Westport, CT: Praeger/Greenwood.
- Kabaservice, G. (2004). *The guardians: Kingman Brewster, his circle, and the rise of the liberal establishment*. New York: Henry Holt.
- Karabel, J. (2006). *The chosen: The hidden history of admission and exclusion at Harvard, Yale, and Princeton*. New York: Mariner.
- Kidder, W. C. (2000). The rise of the Testocracy: An essay on the LSAT: Conventional wisdom and the dismantling of diversity. *Texas Journal of Women and the Law, 9*, 167-217.
- Kidder, W. C. (2003). Law and education: Affirmative action under attack: The struggle for access from Sweatt to Grutter: A history of African American, Latino and American Indian Law School Admissions, 1950-2000. *Harvard BlackLetter Law Journal, 19*, 1-35.
- Kobrin, J. L., Camara, W. J., Milewski, G.B. (2002). *The Utility of the SAT I and SAT II for Admissions Decisions in California and the Nation*. (College Board Report No. 2002-6). New York: College Entrance Examination Board.
- Kogan, N. & Pankove, E. (1974). Long-term predictive validity of divergent thinking tests: Some negative evidence. *Journal of Educational Psychology, 66*, 802-810.
- Leijnen, S. & Gabora, L. (2010). An agent-based simulation of the effectiveness of creative leadership. *Proceedings of the Annual Meeting of the Cognitive Science Society* (pp. 955-960), August 11-14, 2010, Portland, Oregon.
- Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. New York: Farrar, Straus, & Giroux.
- Loehlin, J. C., Lindzey, G., & Spuhler, J. N. (1975). *Race differences in intelligence*. New York: Freeman.

- Lubinski, D., Benbow, C. P., Webb, R. M., & Bleske-Rechek, A. (2006). Tracking exceptional human capital over two decades. *Psychological Science, 17*, 194-199.
- Lynn, R. (2006). *Race differences in intelligence: an evolutionary analysis*. Augusta, GA: Washington Summit.
- McDonough, P. M. (1997). *Choosing colleges: How social class and schools structure opportunity*. Albany, NY: State University of New York Press.
- Merton, R. K. (1968). The Matthew effect in science. *Science, 159*(3810), 56-63.
- Niu, W., & Sternberg, R. J. (2001). Cultural influences on artistic creativity and its evaluation. *International Journal of Psychology, 36*(4), 225-241.
- Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology, 89*, 187-207.
- Plucker, J. A. & Makel, M. C. (2010). Assessment of creativity. *Cambridge Handbook of Creativity*, (pp. 48-73). Cambridge, MA: Cambridge University Press.
- Puccio, G. J., Murdock, M. C., & Mance, M. (2006). *Creative leadership: Skills that drive change*. Thousand Oaks, CA: Sage.
- Renzulli, J.S. (2005). The three-ring definition of giftedness: A developmental model for promoting creative productivity. In R.J. Sternberg & J.E. Davidson (Eds.), *Conceptions of giftedness*, (2nd ed., pp. 246-280). New York: Cambridge University Press.
- Rowe, D. C. (2005). Under the skin: On the impartial treatment of genetic and environmental hypotheses of racial differences. *American Psychologist, 60*(1), 60-70.
- Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law, 11*, 235-294.
- Sacks, P. (2007). *Tearing down the gates: Confronting the class divide in American Education*. Berkeley, CA: University of California Press.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.
- Schmitt, N., Billington, A. Q., Keeney, J., Oswald, F. L., Pleskac, T., Sinha, R., & Zorzie, M. (in press). Prediction of four-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology, 94*, 1470-1497.
- Sedlacek, W. E. (2004). *Beyond the big test: Noncognitive assessment in higher education*. San Francisco: Jossey-Bass.
- Spearman, C. (1927). *The abilities of man*. New York: Macmillan.
- 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. (1997). *Successful intelligence*. New York: Plume.
- Sternberg, R. J. (1998a). Abilities are forms of developing expertise. *Educational Researcher, 27*(3), 11-20.

- Sternberg, R. J. (1998b). A balance theory of wisdom. *Review of General Psychology*, 2, 347–365
- Sternberg, R. J. (1999a). Intelligence as developing expertise. *Contemporary Educational Psychology*, 24, 359–375.
- Sternberg, R. J. (1999b). The theory of successful intelligence. *Review of General Psychology*, 3, 292–316.
- Sternberg, R. J. (2003a). Teaching for successful intelligence: Principles, practices, and outcomes. *Educational and Child Psychology*, 20(2), 6–18.
- Sternberg, R. J. (2003b). *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. (2005a). Accomplishing the goals of affirmative action – with or without affirmative action. *Change*, 37(1), 6–13.
- Sternberg, R. J. (2005b). The theory of successful intelligence. *Interamerican Journal of Psychology*, 39(2), 189–202.
- Sternberg, R. J. (2006). How can we simultaneously enhance both academic excellence and diversity? *College and University*, 81 (1), 17–23.
- Sternberg, R. J. (2007a). Culture, instruction, and assessment. *Comparative Education*, 43 (1), 5–22.
- Sternberg, R. J. (2007b). Finding students who are wise, practical, and creative. *The Chronicle of Higher Education*, 53 (44), B11.
- Sternberg, R. J. (2007c). How higher education can produce the next generation of positive leaders. In M. E. Devlin (Ed.), *Futures Forum 2007* (pp. 33–36). Cambridge, MA: Forum for the Future of Higher Education.
- Sternberg, R. J. (2009). The Rainbow and Kaleidoscope Projects: A new psychological approach to undergraduate admissions. *European Psychologist*, 14, 279–287.
- Sternberg, R. J. (in press). *College admissions for the 21<sup>st</sup> century*. Cambridge, MA: Harvard University Press.
- Sternberg, R. J., Bonney, C. R., & Gabora, L. (in press). Broadening the spectrum of undergraduate admissions. *College and University*.
- Sternberg, R. J., Castejón, J. L., Prieto, M. D., Hautamäki, J., & Grigorenko, E. L. (2001). Confirmatory factor analysis of the Sternberg triarchic abilities test in three international samples: An empirical test of the triarchic theory of intelligence. *European Journal of Psychological Assessment*, 17(1) 1–16.
- Sternberg, R. J., & Coffin L. A. (Winter, 2010). Admitting and developing “new leaders for a changing world.” *New England Journal of Higher Education*, 12–13.
- Sternberg, R. J., & Davidson, J. E. (Eds.) (2005) *Conceptions of giftedness* (2<sup>nd</sup> ed.). New York: Cambridge University Press.
- Sternberg, R. J., Ferrari, M., Clinkenbeard, P. R., & Grigorenko, E. L. (1996). Identification, instruction, and assessment of gifted children: A construct validation of a triarchic model. *Gifted Child Quarterly*, 40, 129–137.
- 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., & Grigorenko E. L. (Eds.). (2002). *The general factor of intelligence: How general is it?* Mahwah, NJ: Lawrence Erlbaum Associates.

- Sternberg, R. J., & Grigorenko, E. L. (2004). WICS: A model for selecting students for nationally competitive scholarships. In A. S. Ilchman, W. F. Ilchman, and M. H. Tolar (Eds.) *The lucky few and the worthy many. Scholarship competitions and the world's future leaders.* (pp. 32–61). Bloomington, IN: Indiana University Press.
- Sternberg, R. J., & Grigorenko, E. L. (2007). *Teaching for successful intelligence* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Corwin.
- Sternberg, R. J., Grigorenko, E. L., & Kidd, K. K. (2005). Intelligence, race, and genetics. *American Psychologist*, *60*(1), 46–59.
- Sternberg, R. J., & the Rainbow Project Collaborators (2005). Augmenting the SAT through assessments of analytical, practical, and creative skills. In W. Camara & E. Kimmel (Eds.). *Choosing students: Higher education admission tools for the 21<sup>st</sup> century* (pp. 159–176). Mahwah, NJ: Lawrence Erlbaum Associates.
- Sternberg, R. J., & The Rainbow Project Collaborators (2006). The Rainbow Project: Enhancing the SAT through assessments of analytical, practical and creative skills. *Intelligence*, *34*, 321-350.
- Sternberg, R. J., The Rainbow Project Collaborators, & University of Michigan Business School Project Collaborators (2004). Theory based university admissions testing for a new millennium. *Educational Psychologist*, *39*(3), 185–198.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago, IL: University of Chicago Press.
- Turkheimer, E., Haley, A., Waldron, M. D'Onofrio, B. & Gottesman, I. I. (2003) Socioeconomic status modifies heritability of IQ in young children. *Psychological Science*, *14*(6):623-628.
- Vernon, P. E. (1971). *The structure of human abilities*. London: Methuen.
- Wallach, M. A. & Kogan, N. (1965). *Modes of thinking in young children*. New York: Holt, Rinehart, & Winston.
- Wallach, M. A. & Wing, C. (1969). *The talented student*. New York: Holt, Rinehart, & Winston.
- Wightman, L. F. (1997). The threat to diversity in legal education: An empirical analysis of the consequences of abandoning race as a factor in law school admission decisions. *New York University Law Review*, *72*(1), 1-53.
- Zwick, R. (2002). [\*Fair game?: The use of standardized admissions tests in higher education\*](#). New York: Routledge.
- Zwick, R., & Green, J. G. (2007). New perspectives on the correlation of SAT scores, high school grades, and socioeconomic factors. *Journal of Educational Measurement*, *44*, 23-45.

---

<sup>1</sup> Liane Gabora's home campus is the University of British Columbia (Okanagan.)