Successful intelligence: A model for testing intelligence beyond IQ tests

CrossMark

Robert J. Sternberg

Cornell University, Department of Human Development, B44 MVR, Ithaca, NY 14853, USA

Abstract   Standard conventional tests only assess a narrow sampling of the abilities required for success in school and in life. In contrast, the augmented theory of successful intelligence asserts that intelligence involves creative skills in producing new ideas, analytical skills in evaluating whether the ideas are good ones, practical skills in putting the ideas into practice and in convincing other people of the value of the ideas, and wisdom-based skills in confirming that one is using one's knowledge and skills to serve a common good. Three projects were created to evaluate the theory with regard to college admissions: First, the Rainbow Project demonstrated that prediction of first-year college academic performance could be increased while simultaneously decreasing differences between ethnic groups on a predictive assessment, in comparison with the Scholastic Aptitude Test (SAT). Second, the Kaleidoscope Project improved prediction of academic and extracurricular performance over SAT scores alone; but the ethnic-group differences usually obtained vanished. Third, the Panorama Project showed the success of similar techniques in a less selective population. The projects demonstrate the application of the augmented theory of successful intelligence in enhancing college and university admissions procedures.

La Inteligencia Exitosa: Un Modelo para evaluar la Inteligencia más allá de los test de Cociente Intelectual (CI)

Resumen   Las pruebas estandarizadas convencionales, evalúan sólo una muestra de las amplias habilidades requeridas para conseguir éxito en la escuela y en la vida. En contraste, la teoría aumentada de la inteligencia exitosa afirma, que la inteligencia implica habilidades creativas en la producción de nuevas ideas, habilidades analíticas para evaluar si las ideas son buenas, habilidades prácticas para implementar las ideas en la práctica y para convencer a otras personas sobre el valor de las ideas, y habilidades basadas en la sabiduría para confirmar que uno

E-mail address: robert.sternberg@cornell.edu

está utilizando sus conocimientos y habilidades para servir a un bien común. Se crearon tres proyectos para evaluar la teoría respecto al acceso a la universidad. En primer lugar, el Proyecto Rainbow (Arco Iris) demostró que la predicción del primer año de rendimiento académico en la universidad se podría aumentar; al mismo tiempo, se puede conseguir la disminución de diferencias entre los grupos étnicos en una evaluación predictiva, en comparación con la prueba de acceso a la Universidad. En segundo lugar, el Proyecto Caleidoscopio, mejoró la predicción del rendimiento académico y extracurricular en comparación con los resultados obtenidos por los alumnos en las pruebas de acceso a la Universidad, pero las diferencias entre los grupos étnicos que se obtienen habitualmente, desaparecieron. En tercer lugar, el Proyecto Panorama mostró el éxito de técnicas similares, en una población menos selectiva (es decir, de menos éxito académico). Los proyectos demuestran que la aplicación de la teoría de la inteligencia exitosa aumentada mejora los procedimientos de acceso a la universidad.

There was a time when even the Model T Ford was a great innovation. But innovations don't stay innovative forever, and today, if you see someone driving a Model T, you are likely to view the car as quaint, antique, passé, or any of a number of other things, but not as innovative or even particularly useful except for generating feelings of nostalgia.

Roughly a century ago, the pioneers of intelligence testing introduced ideas and technological innovations that, for their time, were revolutionary. Since that work by these pioneers, Alfred Binet and Theodore Simon, testing to identify cognitive skills prerequisite to academic and other forms of success has changed relatively little (Binet & Simon, 1916). In contrast, other technologies, such as medical testing, telecommunications, and computation, have changed enormously. No one would want to be tested for cancer with early 20th century technology, or wait to pay for an operator to connect a long-distance call, or look forward to the *future* day when UNIVAC, one of the first computers, could help with their data analysis. If any other technology had stayed about the same for 100 years, people would be amazed. Yet, this retro world is the world in which we live in the field of testing the abilities of the gifted and the not so gifted.

There certainly have been new developments. Joseph Renzulli (2005) and Howard Gardner (1983), in particular, but also others (Sternberg & Davidson, 2005) have proposed new models of identification that have been used to identify gifted children in ways that go beyond conventional IQ testing. But the principal tests used to measure IQ and related abilities have not changed much, whether one is seeking to identify the gifted or those with, say, intellectual disabilities. Moreover, it is not just a matter of measuring ''IQ.'' Other tests that measure largely the same thing as IQ tests, such as SATs and ACTs (Frey & Detterman, 2004), also have changed little over time. Most of the changes in these tests have been cosmetic ones responding to demands from the marketplace, not to scientific advances.

Contemporary standardized tests measure today, as they did earlier, what is called ''general ability,'' which English psychologist Charles Spearman (1904) identified early in the twentieth century. The efforts of my colleagues and I have been addressed toward developing new kinds of ability and achievement tests that assess abilities in broader ways than has been the case in the past. We have sought especially to identify gifted individuals.

We call our framework the augmented theory of successful intelligence, or WICS. This is an acronym for **w**isdom, **i**ntelligence, and **c**reativity, **s**ynthesized (Sternberg, 1997, 2003a, 2005; Sternberg & Grigorenko, 2004). In almost any life pursuit, people need to think (a) *creatively* to generate new and valuable ideas, (b) *analytically* to judge whether their ideas and the ideas of others are worthwhile; and (c) *practically* to implement their ideas and to convince others of the value of those ideas. People also need (d) *wisdom* to help to ensure that their skills are utilized to achieve a common good that balances their own (intrapersonal) interests with other people's (interpersonal) and institutional (extrapersonal) interests over the long term, not just the short term. According to WICS, people can improve in these cognitive skills (Dweck, 1999; Sternberg, 1999, 2003b; Sternberg & Grigorenko, 2007).

On this view, traditional ability tests, originating with those of Binet and Simon (1916) and Spearman (1904), are less than comprehensive because they so strongly focus on analytical (and also memory-based) skills without also assessing creative, practical, and wisdom-based skills. Traditional standardized tests correlate with varied kinds of performances on life tasks (Herrnstein & Murray, 1994; Jensen, 1982; Schmidt & Hunter, 1998), but not at an impressive level of magnitude.

WICS is not the only theory, of course, that proposes abilities beyond general intelligence, something others have done before (Gardner, 2006; Thurstone, 1938). For example, Howard Gardner has argued that there are eight multiple intelligences, not just a single ''intelligence.'' Even theories that specify just one general intelligence generally differentiate abilities levels of cognitive skills hierarchically arranged (Carroll, 1993; Cattell, 1971; Johnson & Bouchard, 2005; Sternberg & Grigorenko, 2002). Where the traditional psychometric theories differ from the more modern ones is in precisely which skills are posited-in what types of cognitive skills are considered sufficiently important to be part of a theory of intelligence-and in how important the skills are considered to be beyond general intelligence (*g*).

School-based assessments of achievement, like standardized tests of academic aptitudes, often emphasize memory-based and analytical skills. For example, the SAT assesses, among other things, vocabulary, analysis of reading passages, and solution of mathematics problems. The memory and analytical skills measured by standardized tests are exactly the ones in which many young people of the American and European middle- and upper middle classes excel. Partly as a result, there is a moderate correlation between test scores and socioeconomic status (Lemann, 1999). The system of selective college admissions, for the most part, is based on tests geared to favor US middle- and upper middle class students, not students of the working class, or of different cultures, who may not have had comparable opportunities (Sternberg, 2004). The current system of standardized tests also is stacked against students from the middle and upper middle classes who learn in nontraditional ways.

On the one hand, then, standardized testing as it now exists can help create equity by contributing to the admissions of students because of their cognitive skills and achievements. But such testing also can contribute to the destruction of equity by giving an advantage to some groups of students over others on bases other than cognitive skills and achievements.

Life success of almost any kind depends on a broader range of skills than is measured by conventional standardized tests. For example, memory and analytical skills may lead to A's in STEM (science-technology-engineering-mathematics) courses, but they probably are not adequate to result in superior research, even if they are relevant, as in evaluating whether one's ideas are worthwhile (Lubinski, Benbow, Webb, & Bleske-Rechek, 2006). Excellent scientific investigators do not just memorize and analyze. They also must creatively generate ideas for theories and/or experiments, analyze whether their ideas are worthwhile, and practically put their ideas into research or practice through funded research and acceptances by refereed professional journals. Ideally, they wisely try to produce some kind of common good with their research. Traditional standardized tests thus may well be a good beginning to identifying the gifted, but, over time, they also appear to have become an end in themselves.

My colleagues and I have been involved in three related projects exploring whether broader scientifically- and 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, and the third, the Panorama Project. These projects are presented in much greater detail elsewhere (Sternberg, 2010, 2012; Sternberg, Bonney, Gabora, Karelitz, & Coffin, 2010; Sternberg & The Rainbow Project Collaborators, 2006). I also describe some other newer projects here.

Projects to broaden the spectrum of admissions

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 (Lemann, 1999). 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 others 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 (Sternberg & Davidson, 2005)? And can one do so in a manner that does not merely echo students' socioeconomic status (Golden, 2006; McDonough, 1997) or IQ (Frey & Detterman, 2004).

Our Rainbow Project (Sternberg & The Rainbow Project Collaborators, 2006) 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 augmented theory of successful intelligence views cognitive skills and achievement as existing on a continuum. On this view, cognitive skills are in large part achieved rather than merely being innate (Sternberg, 1999).

In the Rainbow Project, my collaborators and I collected data from 15 US institutions, including 8 four-year colleges, 5 two-year colleges, and 2 high schools.

A total of 1,013 students participated in the project. Most were in their first year of college or in their last year of high school. The analyses presented here are only for college students because they were the only ones for whom the Rainbow Project team had data relevant to college GPA. The number of college participants upon whose data these analyses are based was 793.

We included standardized test scores and high-school GPA to analyze the predictive validity of tools currently favored in college admission decisions. We also hoped to determine whether we could improve upon the prediction provided by current measures. Students' scores on the SAT were provided by the College Board.

We used as measures of analytical skills the SAT plus multiple-choice analytical items of our own devising. We introduced into the mix three item types. First, inferring meanings of previously unknown words from paragraph contexts assessed students' vocabulary-acquisition skills. Number-series completions assessed students' skill in inductively inferring the next number in a series. Figural-matrix completions assessed students' figural inductive-reasoning skills.

Creative thinking skills were assessed via both multiple-choice items and performance-based items. There were three kinds of multiple-choice items. One kind was verbal analogies preceded by counterfactual premises (e.g., suppose that money fell off trees). Test-takers then had to solve the analogies as though the counterfactual premises of the analogies were true. A second kind of item presented rules for novel number operations. An example is ''flix,'' which involves numerical computations that differ as a function of whether the first of two operands is greater than, equal to, or less than the second. Students employed the

novel number operations to solve the math problems we presented. A third item type involved figural series with one or more transformations; students then had to apply a rule from one series of figures to a new figure with a different appearance, and then complete the new series of figures.

Creative thinking skills also were assessed via open-ended measures. One item type asked test-takers to write short stories. The test-takers selected two titles from a larger set, including unusual topics such as ''The's Sneakers.'' A second item type asked test-takers to tell two short stories orally on the basis of choices from among pictorial collages. The third item type asked test-takers to caption cartoons. Trained raters evaluated the open-ended answers for three characteristics: novelty, quality, and (on a yes or not basis) appropriateness to the task. Multiple judges rated responses for each task. We found satisfactory inter-rater reliabilities for all tasks.

We also used three types of multiple-choice items to assess practical skills. The first type consisted of problems adolescents face in their daily lives. The task of the test-takers was to choose the option best solving each problem. A second problem type confronted test-takers with situations requiring the everyday mathematics (e.g., buying tickets to a baseball game). Test-takers were required to solve mathematical problems based on the situations described. The third task showed test-takers a map (e.g., of an amusement park). Test-takers were asked questions about how navigate a route based on the map that was presented.

Practical skills also were assessed by means of three situational-judgment questionnaires: an Everyday Situational Judgment Questionnaire (Movies), a Common Sense Questionnaire, and a College Life Questionnaire. Each questionnaire measured practical reasoning in an everyday context. The construction and use of such questionnaires is explained elsewhere (Sternberg et al., 2000).

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. The Common Sense Questionnaire contained common problems of kinds encountered in business. One problem was being assigned to collaborate with a disagreeable colleague. The College Life Questionnaire presented situations commonly encountered in college, such as how to prepare for a test or how to write a paper.

Test-takers received varying numbers responses to rate for quality. The test-takers were informed that there was no single correct answer and that the responses furnished for each problem illustrated varied ways in which individuals might react to the situations presented on the test.

Examples of creative tasks in the Rainbow Project were to write very short stories with suggested titles, such as ''3516'' or ''It's Moving Backward.'' In a second task assessing creativity, test-takers were presented with collages of pictures showing individuals involved in a wide range of activities. The test-takers then created and orally told a story that was derived from the pictorial collage.

One type of practical item presented short videos for which test-takers saw scenarios that were incomplete: The test-takers then had to figure out how to response to the scenarios. In one scenario, for example, a college student approaches what appears to be a college professor to ask for a recommendation letter. After a brief conversation with the professor, the student comes to realize that the professor does not know who he is. The video then stops. The test-taker had to indicate how he or she would deal with the 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.

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). 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. The conclusion we reached is that it is important to measure cognitive skills through diverse item formats, not just through a multiple-choice format.

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 our 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.

Thus, the Rainbow assessments substantially increase the level of prediction beyond that resulting from SATs on their own. Our 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.

Studying differences among groups can lead to mistaken conclusions (Hunt & Carlson, 2007). In the Rainbow Project,

my colleagues and I sought to create assessments that would mitigate ethnic group differences. 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 (Camara & Schmidt, 1999; Rowe, 2005; Sternberg, Grigorenko, & Kidd, 2005). There are multiple means available by which investigators can assess group differences in college-admissions test scores. Each means involves a test of size of the effect for ethnic group. We chose two different statistical indices: ω^2 (omega squared) and Cohen's *D*.

What did we 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.

Although the group differences were not eliminated, our 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.

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.

The Kaleidoscope Project

After 30 years, I left my professorship at Yale to become dean of arts and sciences at Tufts University. In collaboration with Dean of Admissions Lee Coffin and other colleagues, I instituted at Tufts Project Kaleidoscope, which represented an operational implementation of the ideas of Rainbow. Kaleidoscope also went beyond Rainbow to incorporate into its assessment the psychological attribute of wisdom (Sternberg, 2007a).

Beginning in 2006 and continuing even to the present day, Tufts 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 WICS—wisdom, analytical and practical intelligence, and creativity synthesized.

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.

Nor was it 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.

We assessed quality of creative and practical thinking in the same way as in the Rainbow Project. We assessed quality of analytical thinking by the organization, logic, and balance of the essay. We 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.

Our goal in Kaleidoscope was not necessarily to replace the SAT, ACT, or other traditional admissions indices such as GPAs and class rank. Instead, our goal was to re-conceptualize applicants in a broader way—in terms of their academic/analytical, creative, practical, and wisdom-based thinking skills. We 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.

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. We therefore were interested in convergent-discriminant validation of our assessments: Would our 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 our 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 our 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 we had sought.

In the first year of Kaleidoscope, the academic credentials (SATs and GPAs) of applicants to Arts and Sciences at Tufts rose slightly. Moreover, we 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.

A fear of some faculty and administrators was that 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.

We 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%.

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.

Other completed projects

When I started as provost and senior vice president of Oklahoma State University, we introduced a project, Panorama, which employed many of the principles of Rainbow and Kaleidoscope to admissions in a large, highly diverse state university. The results were not yet statistically analyzed when I left, but the project was deemed a success by the admissions office in terms of admitting diverse and qualified students who otherwise would not have been admitted.

The principles behind the Rainbow Project apply at other levels of admissions as well. For example, Hedlund, Wilt, Nebel, Ashford, and Sternberg (2006) showed that the ideas of WICS also could be applied to admission to business schools. The goal this project, the University of Michigan Business School Project, was to show 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, our 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, our 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 better coverage of course material (higher content validity) and also reduced obtained ethnic-group differences on the tests (Stemler, Grigorenko, Jarvin, & Sternberg, 2006; Stemler, Sternberg, Grigorenko, Jarvin, & Sharpes, 2009).

Grigorenko and her colleagues wanted to show that a test measuring practical-intellectual skills was relevant to success at the high-school level and could incrementally predict secondary-school success beyond the prediction of a standardized test. Items on the assessment measured skills such as dealing with teachers, with other students, and with homework. Grigorenko and her colleagues found that it was possible to improve prediction of private high school (prep school) performance beyond scores attained on the SSAT (Secondary School Admissions Test) (Grigorenko et al., 2009).

The same principles have been employed in a test for identification of gifted students in elementary school (Chart, Grigorenko, & Sternberg, 2008). In this case, a test, Aurora, was created to predict success in gifted programs at the upper elementary level. The test assesses analytical, creative, and practical skills in the verbal, quantitative, and figural domains.

Current projects

My collaborators and I currently have two projects that are extending the work we have done to admissions for graduate and professional schools. These projects are ongoing so we do not yet have data.

Graduate admissions in the behavioral and brain sciences

A first project is to measure skills relevant for success in graduate school in the behavioral and brain sciences, beyond the skills measured by the GRE (Graduate Record Examination). The assessment we are using has three parts.

First, the test-taker reads about an empirical study a student has conducted, including her hypothesis for why she discovered what she discovered. An example is:

"Eve is interested in studying the effects of taking exams on student performance. She devises an experiment where Group A students are given weekly quizzes and twice-per-semester exams, while Group B students are only given the exams. The results show that students in Group A do better overall than do students in Group B. She explains that weekly quizzes help the students stay on track with the material. *What are some alternative hypotheses regarding why the students who received weekly quizzes perform better than the students who don't?*" An example of an alternative hypothesis would be that students may not have been randomly assigned to groups, and as a result students in Group A may have been more familiar with the subject matter than students in Group B.

A second item type has students read several scenarios that describe a situation as well as a hypothesis. The students are asked to design an experiment for each of the scenarios to test the hypothesis presented. An example is:

"Martin believes that a particular yellow food dye (E104) not only causes hyperactivity in children (as has been shown), but also increases people's creativity. That is, he believes this dye puts people in a state in which they are more creative. How can he test his hypothesis that the dye E104 increases creativity?" An example of a study is to recruit 100 participants. Give half of them in one randomly assigned group a beverage that contains E104, and the other randomly assigned half a beverage with a different dye. Then administer several tests of creative thinking, such as the Torrance Tests of Creative Thinking, to see whether the students who drank the beverage with E104 perform at a higher level. Retest them a week later.

In a third item type, students read several scenarios that describe an experiment that was conducted to test a specific hypothesis. However, each of the experiments is flawed. Students are asked to consider the experimental design and point out the flaw(s). Here is an example:

"We tested the hypothesis that when a salesperson smiles directly at a customer, the individual is more likely to make a sale than when the salesperson fails to smile. Five saleswomen at a bridal shop were instructed to do one of three things while trying to sell a wedding dress to a customer: either to smile directly (in the face of) the customer, smile indirectly (while looking away from) the customer, or have a neutral expression on the face. It was found that smiling directly at customers did increase sales significantly. Fewest wedding dresses were bought in the indirect-smiling condition. It was concluded that salespeople should smile directly into the faces of their customers if they wish to increase their sales effectiveness." In this study, one flaw is that all salespeople and purchasers were women, so it may be that the results would not generalize beyond women; it also may be that the results do not generalize beyond purchase of wedding dresses, a particularly happy occasion for most customers.

All scoring in these items is done by expert raters using rubrics that are provided to them.

Medical school admissions

The MCAT (Medical College Admission Test) measures various types of knowledge and reasoning but it does not assess students' good judgment in actual situations that medical practitioners might find themselves confronting. The idea of our current study is to present potential applicants to medical school with a series of situations that they might encounter as medical practitioners and then to ask them how they would respond to the situations. Most of the situations involve ethical dilemmas. Here are two examples of situations:

1. Doctors sometimes write notes on pads furnished them by pharmaceutical companies with pens also furnished by such companies. Some doctors also may accept free meals, club memberships, subsidized travel, and research funds from such companies. With regard to gifts and subsidies from pharmaceutical companies to doctors, what kinds of guidelines do you think ought to be in place, and why?
2. Mr. Smith, a patient of yours, is clearly dying. There is no hope. On his deathbed, he tells you that he has been burdened for many years by the fact that, between the ages of 35 and 42, he had a mistress whom he saw frequently and subsidized financially. He asks you to tell his wife what he has told you and to tell her that he begs her forgiveness. Mr. Smith has now died. What should you do about his request? Scoring for this project, as for the previous project, is by rubrics.

Conclusion

In conclusion, the augmented theory of successful intelligence provides a theoretical basis for assessing many of the skills needed for college (and other forms of) success. Measures derived from the theory show significant and substantial incremental predictive power, and also increase equity across ethnic groups. If our society were to experience better teaching, with more emphasis on the creative and practical skills needed for success in school and life, the predictive power of WICS assessments might increase

further. Cosmetic changes in assessment during the last century have made relatively little difference to the construct validity of the assessment procedures our society uses. The augmented theory of successful intelligence could provide a new opportunity to increase construct validity and at the same time reduce differences in test performance between ethnic and other groups. It may even be possible to accomplish the goals of affirmative action through tests such as the Rainbow assessments, either as supplements to traditional affirmative-action programs or as replacements for them.

Other modern theories of intelligence, such as those mentioned earlier in the article, 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, Kim, Ramsay, and Gillespie and Neal Schmitt and his colleagues have found biographical data and situational-judgment tests (the latter of which we also used) to provide incremental validity to the SAT (Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004; Schmitt et al., 2009). William Sedlacek has developed non-cognitive measures that appear to have had success in enhancing the university-admissions process (Sedlacek, 2004).

The theory and principles of assessment described in this article can be extended beyond the United States (Sternberg, 2007b). 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.

There is no question but that the methods used in the Rainbow Project, the Kaleidoscope Project, the Panorama 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. But our results show that tests measuring memory and analytical skills tell an incomplete story. To finish the story, we need also to measure creative practical, and wisdom-based skills as well. But these are not the skills that matter, and should not be the only skills we measure (Sternberg, 2003a; Sternberg, Jarvin, & Grigorenko, 2011).

In the field of assessment, our society has been locked into a ''Model-T'' model of how to assess the abilities and achievements of US students. The use of an antiquated model is not obvious because the skin of the testing vehicle has been update to look modern, while the insides of the testing vehicle remain, essentially, the same old same old.

Societal forces have conspired to retain this dated model. First, at least in college admissions, test-takers and not colleges pay for the testing, so it is inexpensive for colleges to keep using the tests. Second, the tests produce exact-sounding numbers so they give the appearance of quantitative precision, even though their validity is only modest to moderate. Third, education today suffers from a great deal of entrenchment-it is hard to get educators to change established practices. Fourth, pressure from accreditors and different levels of government (federal, state, local) to produce high test scores locks schools into existing tests. Fifth, school districts and universities alike compete for higher ratings in the media, and such ratings often are based on test scores. Finally, the people who run educational enterprises themselves generally did well enough on the test to be advanced to their current jobs, so like all of us, they label as ''successful'' others like themselves.

The existing tests are not ''bad.'' Rather, they are incomplete. They measure some of the elements needed for future success but not others. Our society can and should do better by seeking to measure more of the elements needed for success, in order to ensure we do not stifle the success of talented individuals, or provide opportunities to individuals who deserve them only in the most limited ways.

Conflict of interest

The author of this article declares no conflict of interest.

Acknowledgments

References

Binet, A., & Simon, T. (1916). *The development of intelligence in children*. Baltimore: Williams & Wilkins (Originally published in 1905).

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 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.

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.

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.

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.

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. (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.

Johnson, W., & Bouchard, T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence, 33*, 393–416.

Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. New York: Farrar, Straus, & Giroux.

Lubinski, D., Benbow, C. P., Webb, R. M., & Bleske-Rechek, A. (2006). Tracking exceptional human capital over two decades. *Psychological Science, 17*, 194–199.

McDonough, P. M. (1997). *Choosing colleges: How social class and schools structure opportunity*. Albany, NY: State University of New York Press.

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.

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., Vol. 1, 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*, 60–70.

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*, 262–274.

Schmitt, N., Fandre, J., Quinn, A., Oswald, F. L., Pleskac, T. J., Sinha, R., & Zorzie, M. (2009). Prediction of four-year college student performance using cognitive and non-cognitive predictors and impact on demographic status of admitted students. *Journal of Applied Psychology, 94*, 1479–1497.

Sedlacek, W. E. (2004). *Beyond the big test: Noncognitive assessment in higher education*. San Francisco: Jossey-Bass.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal Psychology, 15*, 201–292.

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*, 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. (1999). Intelligence as developing expertise. *Contemporary Educational Psychology, 24*, 359–375.

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

Sternberg, R. J. (2003b). Teaching for successful intelligence: Principles, practices, and outcomes. *Educational and Child Psychology, 20*(2), 6–18.

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

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

Sternberg, R. J. (2007a). 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. (2007b). Culture, instruction, and assessment. *Comparative Education, 43*(1), 5–22.

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

Sternberg, R. J. (2012). College admissions assessments: New techniques for a new millennium. In J. Soares (Ed.), *SAT wars: The case for test-optional college admissions* (pp. 85–103). New York: Teachers College Press.

Sternberg, R. J., Bonney, C. R., Gabora, L., Karelitz, T., & Coffin, L. (2010). Broadening the spectrum of undergraduate admissions. *College and University, 86*(1), 2–17.

Sternberg, R. J., & Davidson, J. E. (Eds.). (2005). *Conceptions of giftedness*. New York: Cambridge University Press.

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, & M. H. Tolar (Eds.), *The lucky few and the worthy many. Scholarship competitions and the world's future leaders* (2nd ed., pp. 32–61). Bloomington, IN: Indiana University Press.

Sternberg, R. J., & Grigorenko, E. L. (2007). *Teaching for successful intelligence* (2nd ed.). Thousand Oaks, CA: Corwin.

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

Sternberg, R. J., Jarvin, L., & Grigorenko, E. L. (2011). *Explorations of the nature of giftedness*. New York: Cambridge University Press.

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

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