

Educational Leadership

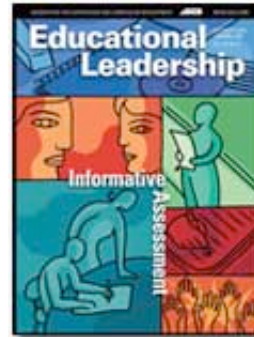
December 2007/January 2008 | Volume 65 | Number 4

Informative Assessment Pages 20-26

Assessing What Matters

Robert J. Sternberg

Worthy assessments should reflect the broader capabilities that students need to thrive in the 21st century.



December 2007/January 2008

My freshman-year introductory psychology course was designed like most courses one finds not just at the college level, but from middle school onward. The main means of teaching was lecture, and the main assessment of performance was a set of tests that measured our recall and basic understanding of the facts taught in the course. I got a C. My professor commented to me, "There is a famous Sternberg in psychology, and it looks like there won't be another one." I got discouraged, left psychology, and came back only when I was failing my introductory course for math majors and decided a C was better than an F.

Thirty-five years later, I became president of the American Psychological Association, which, with a membership of 155,000, is the largest professional organization of psychologists in the world. In some ways, it is the best position one can get in the field of psychology. I cracked to my predecessor that it was ironic that I, who had gotten a C in my introductory course, was now president of the association. He looked me straight in the eye and admitted that he, too, had gotten a C.

This vignette points out in microcosm what may be wrong with the assessments to which we, as a society, have committed ourselves. As a teacher or administrator, how many times have you had to take a multiple-choice or fill-in-the-blank test except perhaps, when you needed to show that you were supposedly qualified for your job? When I look at the skills and concepts I have needed to succeed in my own field, I find a number that are crucial: creativity, common sense, wisdom, ethics, dedication, honesty, teamwork, hard work, knowing how to win and how to lose, a sense of fair play, and lifelong learning. But memorizing books is certainly not one of them.

One can argue, with justification, that one cannot think without content to think with and about. This is indisputable. But when we teach only for facts, rather than for how to go beyond facts, we teach students how to get out of date. For example, the facts that I learned in my introductory psychology course matter little today. An introductory text today contains almost entirely different facts. I know: I am the author of one of those textbooks (Sternberg, 2004). Other fields, such as the hard sciences, political science, economics, and so forth, change at least as rapidly. Even the humanities change: A set of classic works remains, but the interpretations—and even what constitutes such interpretations—change.

So what should we assess? We should assess what students need to become active and engaged

citizens of the world in which they will live—in a sense, what it takes to be "expert" citizens. Oddly enough, a lot of models can prepare students for the roles they will play in their world. Traditional schooling just does not happen to be one of them. We should also assess in ways that can help students develop the skills they need for success in school and life.

Consider students on an athletic team. They learn declarative knowledge about the sport. But learning the rules of the game will no more help them in playing the game than memorizing a book of rules on driving will help someone drive. The students also need to learn how to play the sport.

But the most important skills they learn have nothing to do with one sport or another. These skills are very much like those I mentioned previously: dedication, honesty, teamwork, common sense, and the wisdom to distinguish right from wrong. Athletics is not the only model for such learning. Consider the members of an orchestra or of a dance ensemble. They, too, must learn to work together and must develop similar skills.

How might assessments better reflect the kinds of skills that matter—not just in school, but also in life beyond school? This is a question that we in the Center for the Psychology of Abilities, Competencies, and Expertise, formerly at Yale and now at Tufts University, have posed for ourselves. It is a challenge that we have, to some extent, taken as our life work.

Assessing for WICS

The model that underlies our assessments is called WICS, which is an acronym for *wisdom, intelligence, and creativity, synthesized* (Sternberg, 2003). The basic idea underlying this model is that active and engaged citizenship and especially leadership require individuals to have (1) a creative vision for how they intend to make the world a better place, not just for themselves, but also for their family, friends, colleagues, and others; (2) the analytical intellectual skills to be able to explain why their vision, and that of others, is a good one; (3) the practical intellectual skills to be able to execute their vision and persuade others of its value; and (4) the wisdom to ensure that their ideas represent a common good, not just their own interests or those of their friends or family. Can we apply this model to assessments that can be used in schools? We have done a variety of projects suggesting that we can.

The Successful Intelligence Model

Some of our earlier projects were based on the predecessor of WICS—the model of successful intelligence (Sternberg, 1997). The programs in this model were designed to determine whether we could teach and assess students for memory and for analytical, creative, and practical achievement in the context of any academic subject at any grade level. At that point, wisdom was not separated from practical skills, although it is distinguishable from them. Wisdom involves using academic and practical intelligence, as well as creativity and knowledge, for a common good. If, for example, a used-car salesman convinces customers to buy bad cars, he could be high in practical (or emotional) intelligence without being wise.

As an example, in social studies, we might assess understanding of the American Civil War by asking such questions as (1) Compare and contrast the Civil War and the American Revolution (*analytical*); (2) What might the United States be like today if the Civil War had not taken place (*creative*)? (3) How has the Civil War affected, even indirectly, the kinds of rights that people have today (*practical*)? and (4) Are wars ever justified (*wisdom*)?

In English, we might assess understanding of a novel such as *The Adventures of Tom Sawyer* by asking (1) How was the childhood of Tom Sawyer similar to and different from your own childhood (*analytical*)? (2) Write an alternate ending to the story (*creative*); (3) What

techniques did Tom Sawyer use to persuade his friends to whitewash Aunt Polly's fence (*practical*)? and (4) Is it ever justified to use such techniques of persuasion to make people do things they do not really want to do (*wisdom*)?

In science, we might ask (1) What is the evidence suggesting that global warming is taking place (*analytical*)? (2) What do you think the world will be like in 200 years if global warming continues at its present rate (*creative*)? (3) What can you, personally, do to help slow down global warming (*practical*)? and (4) What responsibility do we have, if any, to future generations to act on global warming now before it gets much worse (*wisdom*)?

In mathematics, we might ask (1) What is the interest after six months on a loan of \$4,000 at 4 percent annually (*analytical*)? (2) Create a mathematical problem involving interest on a loan (*creative*); (3) How would you invest \$4,000 to maximize your rate of return without risking more than 10 percent of the principal (*practical*)? and (4) Why do states set maximum rates of interest that lenders can charge, and should they do so (*wisdom*)?

We have found in studies of reading, social studies, science, and mathematics at a variety of grade levels that teaching for analytical, creative, and practical thinking, as well as for memory, boosts achievement on tests that measure achievement broadly, across subject-matter areas and grade levels (see Grigorenko, Jarvin, & Sternberg, 2002; Sternberg, Grigorenko, Ferrari, & Clinkenbeard, 1999; Sternberg, Torff, & Grigorenko, 1998). Interestingly, even when students are assessed solely for memory, they perform better when taught broadly than when taught just for memory. This is because broader teaching enables students to capitalize on their strengths and correct or compensate for their weaknesses in learning. For example, broader teaching might involve encouraging students who are more visually oriented and less numerically oriented to draw a diagram to help them visualize and solve an algebra problem. Students who are more numerically oriented might proceed directly to constructing a set of equations.

Assessing Creative and Practical Thinking

In our society, a problem with teaching and assessing more broadly is that the kinds of standardized assessments we currently use are quite narrow. For example, the SAT Reasoning Test and the SAT Subject Tests assess primarily remembered knowledge and analytical skills applied to this knowledge. Creativity, practical thinking, and wisdom are assessed minimally or, more likely, not at all. Is there any hope that our society can transport some of these skills to high-stakes assessments?

My collaborators and I decided to find out. In one study, the Rainbow Project, we designed tests of creative and practical thinking that could supplement tests like the SAT Reasoning Test, which measures analytical skills in the verbal and mathematical domains. We tested 1,013 high school students and college freshmen from 15 different schools. We posed analytical questions much like those traditionally found on standardized tests. But we also asked the students to answer creative and practical questions.

The creative tests required the students to stretch their imaginations. For example, they might be asked to write a creative story with a title like *The Octopus's Sneakers* or *3821*. Or they might be shown a collage of pictures, such as of musicians or athletes, and be asked to tell a story about the collage. Or they might be asked to caption an untitled comic strip.

The practical tests required the students to solve everyday problems. Some tests were presented verbally; others, through videos. For example, students might see a movie showing a student about to ask a professor for a letter of recommendation, but also showing the blank look on the professor's face, indicating that he did not know who the student was. The task would be to decide what the student should do. Or students might see a video that shows a group of

friends trying to figure out how to move a large bed up a winding staircase.

There were three crucial findings (Sternberg & the Rainbow Project Collaborators, 2006). First, in addition to the information that the tests provided about students' creative and practical thinking capabilities, we learned something important about multiple-choice problem solving: Multiple-choice tests, no matter what they were supposed to measure, clustered together. Students who were better at one multiple-choice test tended to be better at others as well. This result suggested that using multiple-choice tests consistently tends to benefit some students and not others.

Second, we discovered that using broader tests for college admissions can enhance academic excellence. When compared with using SAT scores alone for predicting freshman-year grades, using these broader tests enabled us to double the accuracy of that prediction. Compared with the predictive value of SAT scores and high school grade point average combined, we increased the accuracy of prediction by about 50 percent. In other words, our assessments were not quixotic ventures into esoteric realms. On the contrary, they enhanced our ability to predict who would be more, as opposed to less, successful in college, at least from an academic point of view.

Third, we discovered that we could substantially reduce ethnic group differences with the tests. In other words, using such tests could increase the proportion of ethnic minorities admitted to selective colleges. The tests would not compromise academic excellence, but actually enhance it. Because different ethnic groups have different conceptions of what intelligence is (Sternberg, 2006), they tend to socialize their children to be intelligent in different ways. For example, on our tests, American Indians, on average, performed lower than most other groups on analytical assessments. But on oral storytelling, they had the highest average scores. Different groups excel, on average, in different ways; giving them a chance to show how they excel enables them to show that they can succeed.

Tests like the Rainbow Assessment do not benefit only members of ethnic minority groups. Many students who come from the majority group, and even from well-off homes, learn best in ways that are different from those assessed by conventional standardized tests. Our tests help identify such students.

Increasing Quality and Diversity

It is one thing to have a successful research project, and another actually to implement the procedures in a highstakes situation. We have had the opportunity to do so.

In 2005, I moved from Yale University, where I was the lead collaborator in the Rainbow Project, to Tufts University, where I became dean of the School of Arts and Sciences. Tufts University, under the leadership of its president, Lawrence Bacow, has strongly emphasized the role of active citizenship in education. So it seemed like an ideal setting to put into practice some of the ideas from the Rainbow Project. In collaboration with Linda Abriola, dean of the School of Engineering, and Lee Coffin, dean of admissions, I instituted Project Kaleidoscope, which implements the ideas of Rainbow but goes beyond that project to include in its assessments the construct of wisdom.

On the 2006–07 application for all of the more than 15,000 students applying to the schools of Arts, Sciences, and Engineering at Tufts, we placed questions designed to assess WICS (Sternberg, 2007). Whereas the Rainbow Project was a separate high-stakes test administered with a proctor, the Kaleidoscope Project was a section of the Tufts college application. The advantage of the Kaleidoscope Project is that it got us away from the high-stakes testing situation in which students must answer complex questions in very short amounts of time under

incredible pressure. The section was optional this past year, and students were encouraged to answer just a single question.

For example, a creative question asked students to write stories with titles like "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 turned 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 to adopt an unpopular idea. A wisdom question asked students how they might apply a passion they had toward the common good.

We now have the results of our first year of implementation, and they are promising. Some stakeholders were afraid that the number of applications would go down; instead, they went up slightly. More notable, the quality of applicants rose substantially. There were 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, decided not to bother to apply. Other stakeholders were afraid that average SAT scores might plummet. Instead, they went up. This is because the new assessments are not negatively correlated with SAT scores. Rather, they are not much correlated at all.

So adopting these new methods does not result in admitting less-qualified applicants. Rather, admitted applicants are *more* qualified, but in a broader way. Moreover, after several years in which the number of applications by underrepresented minorities remained relatively flat, this year they increased substantially. In the end, we admitted 30 percent more black students than the year before and 15 percent more Hispanics. Our results, like those of the Rainbow Project, showed that it is possible to increase academic quality and diversity simultaneously and to do so for an entire undergraduate class at a major university. Most important, we sent a message to students, parents, high school guidance counselors, and others that we believe there is more to a person than the narrow spectrum of skills assessed by standardized tests and that we can assess these broader skills in a quantifiable way.

Such projects can be done at any level. We designed an admissions test for a well-known private school, which showed results for a whole class that were comparable to those for the Rainbow Project. We also did a project in a large business school and showed that we could increase the accuracy of prediction and decrease both gender and ethnic group differences in admissions (Hedlund, Wilt, Nebel, Ashford, & Sternberg, 2006). We are currently developing a comparable test for middle school students (Chart, Grigorenko, & Sternberg, in press).

One might wonder how to assess responses to questions that seem so subjective. The answer is through well-developed rubrics. For example, we assess analytical responses on the basis of the extent to which they are analytically sound, balanced, logical, and organized. We assess creative responses on the basis of how original and compelling they are, as well as on the basis of their appropriateness to the task presented. We assess practical responses on the basis of how feasible they are with respect to time, place, and human and material resources. We assess wisdom-based responses on the extent to which they promote the common good by balancing individual interests with others' larger interests, over the long and short terms, through the infusion of positive (prosocial) values.

Promoting Wisdom

Perhaps conventional assessments met the cognitive demands placed on students 100 years ago. They do not meet the cognitive demands of the world today. Active and engaged citizens must be creatively flexible, responding to rapid changes in the environment; able to think critically

about what they are told in the media, whether by newscasters, politicians, advertisers, or scientists; able to execute their ideas and persuade others of their value; and, most of all, able to use their knowledge wisely in ways that avoid the horrors of bad leadership, as we have seen in scandals involving Enron, Arthur Andersen, Tyco, Clearstream, and innumerable other organizations.

It may be a hard sell to teach and assess for wisdom. However, wisdom is the most important and yet most neglected aspect of education today (Sternberg, 2001a, 2001b). We have seen in failed leaders the enormous costs of having leaders who are knowledgeable and intelligent—who have "good degrees" from prestigious schools—yet who are unwise. They tend to commit several serious cognitive fallacies. They are (1) *unrealistically optimistic*, believing that anything they do will turn out well because they are so brilliant; (2) *egocentric*, believing that the world revolves around them; (3) *falsely omniscient*, failing to learn from experience because they believe they know everything; (4) *falsely omnipotent*, believing that they are all-powerful by virtue of their superior skills or education; (5) *falsely invulnerable*, believing they can get away with almost anything because they are so clever; and (6) *ethically disengaged*, believing that ethical principles apply only to lesser mortals. In my view, much of what is wrong in the world today stems from people who are simultaneously smart and foolish.

Four caveats are in order here. First, my work on WICS and successful intelligence is not the only theory on the basis of which we might create new, broader assessments. Howard Gardner's (1999) theory of multiple intelligences provides another basis for such assessments, and other theories could be used as well. Second, the assessments do not measure all the skills required for success in everyday life. For example, although I assess teamwork in the courses I teach, the assessments I have described do not measure this skill, at least not directly. Third, the assessments have not been scaled up for use on a statewide or national basis. Doing so would no doubt present new challenges. Fourth, expanded assessments cost more time and money. But when we consider the benefits of opening up possibilities and hope to diverse students who learn and think in a variety of ways—whatever their gender or ethnic background—the costs may be relatively small.

Worthy and Wise

There is another issue we need to face. Traditional assessments provide little help to students in learning how to capitalize on strengths and compensate for or correct weaknesses. They measure only narrow bands of skills. Broader tests can give broader ranges of scores and help students see where they have mastery and where they need to improve. Teachers, in turn, can teach in ways that help students acquire the skills they need to succeed in school and life (Sternberg & Grigorenko, 2000, 2007). From this point of view, instruction and assessment are two sides of the same coin rather than two different coins. Assessment drives instruction.

So let's create assessments that are worthy of such a role. To prepare students for a world in which political, economic, social, and even climatic contexts are rapidly changing, we must focus on more than just facts and figures. Our society needs citizens and leaders who are not just memorizers and who are more than just analytically adept. We need people who are creative, practical, and, especially, wise.

References

Chart, H., Grigorenko, E. L., & Sternberg, R. J. (in press). The Aurora Battery: Toward better identification of giftedness. In C. Callahan & J. Plucker (Eds.), *What the research says about: An encyclopedia of research on gifted education*. Waco, TX: Prufrock Press.

Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. New York: Basic Books.

Grigorenko, E. L., Jarvin, L., & Sternberg, R. J. (2002). School-based tests of the triarchic theory of intelligence: Three settings, three samples, three syllabi. *Contemporary Educational Psychology, 27*, 167–208.

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.

Sternberg, R. J. (1997). What does it mean to be smart? *Educational Leadership, 54*(6), 20–24.

Sternberg, R. J. (2001a). Why schools should teach for wisdom: The balance theory of wisdom in educational settings. *Educational Psychologist, 36*(4), 227–245.

Sternberg, R. J. (2001b). Wisdom and education. *Perspectives in Education, 19*(4), 1–16.

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

Sternberg, R. J. (2004). *Psychology* (4th ed.). Belmont, CA: Wadsworth.

Sternberg, R. J. (2006). Recognizing neglected strengths. *Educational Leadership, 64*(1), 30–35.

Sternberg, R. J. (2007). Finding students who are wise, practical, and creative. *The Chronicle of Higher Education, 53*(44), B11.

Sternberg, R. J., & Grigorenko, E. L. (2000). *Teaching for successful intelligence*. Arlington Heights, IL: SkyLight.

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

Sternberg, R. J., Grigorenko, E. L., Ferrari, M., & Clinkenbeard, P. (1999). A triarchic analysis of an aptitude-treatment interaction. *European Journal of Psychological Assessment, 15*(1), 1–11.

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

Sternberg, R. J., Torff, B., & Grigorenko, E. L. (1998). Teaching for successful intelligence raises school achievement. *Phi Delta Kappan, 79*, 667–669.

Robert J. Sternberg is Dean of the School of Arts and Sciences, Professor of Psychology, and Adjunct Professor of Education at Tufts University, Medford Massachusetts; Robert.Sternberg@tufts.edu.

Copyright © 2007 by Association for Supervision and Curriculum Development

© Copyright ASCD. All rights reserved.