

**Toxic testing: It's time to reflect upon our current testing practices**

Betsy Gunzelmann

Southern New Hampshire University

Over many years, both as an educator and as a psychologist, I have been involved with assessing students' knowledge and skills. Since my early years diagnosing students with learning disabilities and other disorders, my viewpoint on the usefulness and effectiveness of many testing approaches has taken a 180-degree turn. In fact, I see many approaches as inaccurate and even harmful for many students.

Take Scott, for example<sup>1</sup>. Scott is an easygoing thirteen-year-old who throughout his school years has been a delight, but oftentimes a puzzle, to his parents and teachers. He is a conscientious student who overall still likes school. He does very well on classroom work and easily follows the lectures and activities, although at times he gets his teachers off-track with his insatiable curiosity and unique approach to problems. He seems to know a lot about many subjects and is inquisitive about most topics.

Scott has a few close friends, no diagnosed learning difficulties, and several intense interests. Math, language arts, and Spanish are his favorite subjects. So what is the problem? one might ask. Well, Scott is one of those bright, creative students who underachieve on standardized, multiple-choice-format tests, despite honor-roll grades and accelerated classes.

Many students like Scott know more than they demonstrate on standardized testing. Very few countries use standardized testing with children before the age of sixteen. But in the United States we use such tests with very young children, even though we know that the practice runs counter to research. Furthermore, very few countries use multiple-choice formats with children of any age (Kohn 2000).

So what's going on? There are many hidden problems with our current approach to testing, and it's high time that we understand the issues behind the obsession with standardized testing in our country.

## **History**

To understand how we got into this testing dilemma, it's necessary to take a step back and look briefly at specific historical precedents in psychology and education. Testing and assessment procedures have been a part of civilized society for centuries, and individual differences have been recognized since the dawn of history. The first tests were designed by

---

<sup>1</sup> The case histories are composites of numerous individuals I have known over the past twenty years. The stories reflect many of the students that psychologists and educators encounter on a daily basis.

the ancient Chinese around 2200 B.C. Plato and Aristotle wrote on individual differences some 2,500 years ago. Many of the early tests used were oral and subject to certain biases (Aiken 2000).

By the late 1800s and early 1900s, testing was beginning to take on a more critical role. In 1904 Binet and Simon were asked to develop the first intelligence test to weed out children who would not benefit from traditional schools (Aiken 2000). Certainly this approach does not meet our needs today, when all children are entitled to an education.

Early psychological theory was closely related to philosophy and understanding the world through a qualitative methodological approach. Then, in Germany during the late 1800s, Wilhelm Wundt started developing research approaches that would allow quantifiable results. In the United States during the late 1940s and 1950s, the behaviorist B. E Skinner and numerous other psychologists and educators became caught up in the need to make everything measurable. Unfortunately, the outcome was that people put more credence in these numbers than was healthy and forgot about the importance of measuring both qualitatively and quantitatively.

Significant historical and political developments also prompted changes in testing needs. For example, both world wars brought about increased calls for innovative approaches to testing many recruits in a short time. Then came the race to get the first man in space and the first man on the moon: the domino effect resulted in a frenzied attempt to increase student learning in math and science. One might cite this as the beginning of an academic Olympics among the industrialized countries. Thus the big business of testing was born along with an increased fervor for competition in academics.

These earlier attempts at testing seem to have established a mold that has been difficult to break. The pattern seems to address the style needs and thinking strategies of many students, but it does not adequately address the needs of all. In fact, I think it is fair to question whether we are actually measuring the true abilities of any student!

### **Current Understanding**

Psychologists and educators know that it is wrong to make decisions based upon a single test score and that decisions should reflect a balanced, complete understanding of each child. Numbers and scores can be very misleading if we don't consider the whole picture,

something that means using both a qualitative and quantitative approach. Yet due to economic, time, and political pressures, psychologists and educators are forced to rely more and more heavily on solely quantitative methods, and many have been deceived into believing that numbers tell the whole story.

Across the country we see continued movement toward more accountability, increased use of standardized tests, and high-stakes testing. Along with these trends come teaching to the test, test anxiety, lowered self-esteem, misunderstanding of children, and missed opportunities for many. Dr. David Elkind, the author of *The Hurried Child* (1989), believes that our current testing obsession is a factor behind the dynamics of our hurrying schools. Administrators, under pressure to demonstrate student learning, are therefore teaching concepts at earlier and earlier ages. The result is no greater knowledge but added pressure for children to measure up and to hurry their learning.

***“There are plenty of kids who think deeply and score well on tests.  
There are also plenty of students who do neither.” –Alfie Kohn***

Deborah Meier (2002) believes that the increased use of standardized tests actually undermines student achievement and increases distrust of teachers, students, and our own judgments. The misunderstanding of testing develops toxic conditions for everyone affected by test scores: students, teachers, parents, administrators, and the entire school system and community. We know from research that no one test can determine a student’s ability or achievement. Nor is there a test that can measure a teacher’s or school system’s effectiveness. To think otherwise is a flagrant misuse of testing.

Much of the drive toward greater accountability is fueled by political platforms. But well-meaning politicians, untrained in the art and science of testing, are influenced by the huge testing industry. Our children’s education is too important to leave assessment decisions in the hands of those who do not comprehend the underlying issues.

### **Problems with Standardized Testing**

Traditional tests attempt to show what a child does not know or what is wrong or deficient about a child’s abilities, rather than what is valued and unique about the child’s particular way of learning, coping, reasoning, or problem solving. Test developers are

looking at assessment too narrowly. We need to break out of the mold of traditional assessment and develop assessment procedures that value the uniqueness of each individual.

Traditional testing is, at best, a selection of test items, which may or may not be relevant to the curriculum to which the student has been exposed and is always subject to many forms of bias, including cultural, gender, socioeconomic, and learning-preference bias. Bias leads to assessment discrimination against many students, including creative thinkers; students with learning differences; students with a preferred learning modality; boys (due to gender differences); students from various ethnic and cultural backgrounds; and many students from lower socioeconomic backgrounds. That is a lot of children!

With so many children at risk, why are we reliant on traditional testing approaches? As I stated earlier, *testing is a big business*, and the testing manuals advertise questionable advantages. For example, traditional standardized testing allows:

1. *Standard practices and scoring.* The tests are given to all students in the same way and scored the same way. Standardizing the process, however, does not eliminate subjectivity. We are still making judgments, but in the case of most standardized testing instruments we are using very little information to make judgments upon individuals. This is a toxic situation for many students
2. *Comparing students.* When students are rank-ordered, the process ensures that half will always be below average.

Such practices give the illusion of being scientific, yet educators know from the work of Brazelton and Greenspan (2000) that they must meet individual needs in order for children to learn and thrive. A one-size-fits-all approach to testing does not address individual needs. Many children have learning difficulties or just learning-style differences. Most teachers do a good job of addressing these preferences in their teaching, and then testing ignores the individual needs of most children. We expect them all to perform using one format. This is testing-preference discrimination. Students should be allowed to demonstrate their competence in ways that show what they really know and are capable of doing.

Thus kids like Scott with unique, idiosyncratic responses to test items are penalized. Their creative, deeper-thinking approaches can actually handicap them on standardized tests. For example, a creative student may come up with correct answers that may not be what the test designers score as correct.

Alfie Kohn (2000) gets this point across succinctly. He believes there is a correlation between high scores on standardized tests and relatively shallow thinking and that these tests are geared to a different, less-sophisticated kind of knowledge (p. 9). Kohn goes on to say,

There are plenty of kids who think deeply and score well on tests. There are also plenty of students who do neither. But as a rule, good standardized test results are more likely to go hand in hand with a shallow approach to learning rather than with deep understanding. (p. 10)

Deborah Meier (2002) concurs that deeper and subtler thought is often an impediment to scoring high on such tests. Meier and Kohn are not alone in their beliefs about deep thinkers. Even in 1962, Banesh Hoffmann, in the classic book *The Tyranny of Testing*, demonstrates that these tests penalize the finer mind.

He [the deep thinker] would see more in a question than his superficial competitors would ever dream was in it, and would expend more time and mental energy than they in answering it. That is the way his mind works. That is, indeed, his special merit. But the multiple choice tests are concerned solely with the candidate's choice of answer, not with his reason for his choice. Thus they ignore the elusive yet crucial thing we call quality. (p. 99)

Furthermore, Hoffmann states: "Multiple choice format also penalizes the creative student. Students who can imagine several possible correct answers, and may think the most obvious answer could not be the correct answer" (p. 101).

The effects of socioeconomic levels and race on test scores have been well documented in the literature. However, there are lesser-known issues of gender impacting standardized test scores. Dr. William Pollack (1998) believes that most schools fail boys if the environment is not conducive to the way boys learn. Many schools are not, and we are seeing a decline in boys' test scores as a result.

Beginning in kindergarten, boys are expected to achieve a standard that favors girls. This standard is reflected in traditional assessment strategies as well. Left-brained cognitive skills such as speaking, reading, and writing abilities tend to develop more slowly in boys, yet both genders are expected to show competence in these areas at the same ages.

Conversely, until the curriculum was changed to meet the needs of girls, girls used to fall behind boys on standardized tests (Connell and Gunzelmann 2004).

W. J. Popham (1999) believes that educational quality is being measured by the wrong yardstick, and the evaluations are therefore apt to be in error. He also believes that most educators as well as parents do not really understand why standardized tests often provide misleading estimates of learning and of a school's effectiveness. Yet many of these tests are being used for high-stakes decisions, including student promotion or retention in grades, graduation, and acceptance into certain schools, as well as judgments and punishments for the teacher--purposes for which the tests were not designed and cannot really accomplish!

All this has developed from tests that we know unfairly discriminate against a variety of students; show only a limited sample of behavior; presume similarity of educational content across classrooms; ignore individualizing ideas such as progressivism and constructivism; make teachers and administrators narrow their curriculum to the test content; and require teachers to focus on test-taking skills, thus forfeiting valuable instruction time. Furthermore, we are wasting taxpayer money on these tests rather than using the funds for educational materials that would enrich all students. Most important, we are not obtaining an accurate picture of many students who may suffer humiliation and serious consequences from low scores.

It puzzles me how our country prizes and protects the rights of the individual, yet our approach to educational testing forces everyone to demonstrate learning in the same way. Because we are all individuals with different achievements, learning styles, backgrounds, and response styles, only a variety of testing formats will allow test takers to accurately demonstrate their learning.

Of course we need high educational standards, but we need to be more reflective about our purposes for testing. Such purposes might include the need to pinpoint learning problems accurately in order to design appropriate educational programs, to improve the learning of all students, and to demonstrate that the children within our classrooms and schools are learning. In order to accomplish those goals of testing we need to develop more accurate assessment tools that do not have toxic side effects.

***“[The deep thinker] would see more in a question than his superficial competitors would ever dream was in it, and would expend more time and mental energy than they in answering it.”—Banesh Hoffmann***

We are overusing and misusing a fallible method of assessment. We can no longer afford to rely on a limited repertoire of assessment approaches. In order to break out of our current obsessive pattern of testing, we need innovative, motivated thinkers who know children well and realize the limitations of traditional tests--individuals who can develop and fine-tune approaches that measure learning over time, not just take a snapshot of a specific behavior at a specific time.

### **Toward a Workable Solution**

Alternative assessment approaches can portray each student’s unique abilities and learning styles. For example, many teachers have been using a form of portfolio assessment for some time (some more effectively than others). Portfolio assessment can be time-consuming, yet it has the added benefit of helping students to take responsibility for their learning and pride in their accomplishments. Used well, portfolio assessment can demonstrate student learning as well as strengths and weaknesses. With further refinement, this approach (along with other qualitative and quantitative approaches) could be used to compare students’ abilities and demonstrate the effectiveness of teachers’ performance as well.

### *The Case of Noel*

The case of Noel demonstrates the power of portfolio assessment. Noel experienced difficulty in her early grades in elementary school. Reading and writing did not come easily for her. Diagnosed with a learning disability during fourth grade, she functioned below her peers in language arts despite high intelligence scores. Noel began to think of herself as less smart than her peers; her self-esteem and self-confidence began to erode. Test after test, year after year, her scores showed only her weaknesses.

Noel developed a pessimistic outlook toward her future and felt trapped by a misleading approach to understanding her knowledge and skills. Not until Noel reached college did she finally begin to understand her strengths and value her abilities. Through



portfolio assessment, she demonstrated her skills to herself as well as to her professors. She became a confident and competent young woman, graduating from college with honors, and she is now applying to graduate school with, strong, well-earned recommendations from her professors.

Noel was unique because she didn't give up. I suspect that many individuals have not persevered after test scores put roadblocks in their way. What a waste it is when human potential goes unrecognized, or even more sadly is misunderstood!

### *The Case of Suzie*

Suzie had always been an excellent student who performed well on all standardized tests administered throughout her elementary and secondary school years. While in college she continued on this path, clearly demonstrating her abilities as a scholar. Yet with all the praise and glory, Suzie had focused too narrowly on scores and competition. She did not understand all her numerous strengths. So even in Suzie's case, testing did not reveal a complete picture of what this young woman might become.

When Suzie became involved with an evaluation process that required her to focus on her strengths and weaknesses, likes and dislikes, she discovered new talents and a balance to her life that allowed more focused career goals to emerge. Suzie was involved with experiential learning in her chosen field, trying out her knowledge and skills. This approach also allowed outside, objective professionals as well as her professors to evaluate her abilities. Suzie took responsibility for her learning and discovered artistic talents, strong interpersonal skills, and a desire and ability to help others. Focusing only on scores, rank-ordering each student, and concentrating on what a child does not know make it easy to lose sight of the individual and what a person can become. Portfolio assessment is just one approach we should be looking at more seriously for use with all students, not just the severely handicapped.

### **Future Directions**

Additional research must be conducted on alternative assessment techniques. Advances in technology afford us many unique opportunities to meet the needs of all

students. For example, Scott, whose case was presented earlier, discovered that he did much better on computerized versions of traditional tests.

Other technological advances have allowed us to use e-portfolios with many added benefits. Storage of materials is much easier; many reviewers can assess students' material online for a more complete evaluation; and students can prepare and present their own portfolios, thus cutting the cost and time required.

Most professionals in psychology, social work, medicine, and education strive to do good work with children. Once in a while, it is a good idea to step back and reflect upon the ramifications of our work so that we can improve our future efforts. We know that:

- Assessment should be undertaken for the right reasons.
- Assessment should be helpful to the students and teachers.
- Assessment must be accurate.
- Assessment should not cause harm.

Relying too heavily on traditional standardized assessment harms many individuals and does not yield accurate understanding. Therefore, more precise assessment requires a combination of qualitative and quantitative approaches. There is no doubt that both methodologies require more research and refinement, but we must not allow judgments and decisions to be made on our children without a complete picture. We have underutilized alternative approaches that can yield a more complete and accurate picture of all students, while they are ultimately affirming and motivating ... in a word, nontoxic!

## References

- Aiken, L. R. 2000. *Psychological Testing and Assessment*. Boston: Allyn & Bacon.
- Brazelton, T. B., and Greenspan, S. 2000. *The Irreducible Needs of Children: What Every Child Must Have to Grow, Learn, and Flourish*. New York: Perseus Books.
- Connell, D., and Gunzelmann, B. 2004. "The New Gender Gap: Why Are So Many Boys Floundering and Girls Soaring?" *Instructor* 113 (6): 14-17.
- Elkind, D. 1988. *The Hurried Child*. Reading, Mass.: Perseus.
- Hoffmann, B. 1962. *The Tyranny of Testing*. New York: Macmillan.

- Kohn, A. 2000. *The Case against Standardized Testing: Raising Scores, Ruining the Schools*.  
Portsmouth, N.H.: Heinemann.
- Meier, D. 2002. *In Schools We Trust. Creating Communities of Learning in an Era of  
Testing*. Boston: Beacon Press.
- Pollack, W 1998. *Real Boys: Rescuing Our Sons from the Myth of Boyhood*. New York:  
Henry Holt.
- Popham, W J. 1999. "Why Standardized Tests Don't Measure Educational Quality." *Using  
Standards and Assessments* 56 (6): 8-15.