PUTTING STANDARDIZED TESTING TO THE TEST

BY JERRY KEA

As public officials and educators seek to improve and strengthen the American school system, a generous use of standardized testing (IQ, achievement, placement, college and graduate school entrance exams) remains a centerpiece of student evaluation.

It would appear that American confidence in standardized tests has never been greater—they are used to employ and dismiss teachers, evaluate schools, confer scholarships, allocate money, and sort and process students from kindergarten through graduate school and into the job market. Ironically, while being used for so many purposes, these all-purpose assessment devices continue to receive a good deal of criticism.

This article will discuss the history, design, and structure of standardized tests, their status in schools, and alternatives to their use for teachers who desire a clearer philosophical focus and more responsible evaluation of Seventh-day Adventist youth.

Testing and grades set up many students for failure in a system that forced all children to attend school, entering at progressively younger ages, and to learn more at an increasingly faster pace, regardless of their ability, need, or interests.

History

Until fairly recently, education occurred in the home or workplace. Here the student's success was determined by performance. To be a farmer, you had to grow wheat; to be a hunter, you had to kill game. If you attended a formal school or studied under a tutor, you were evaluated on whether you successfully memorized facts or mastered certain skills. Examinations were competency-based and used to determine the student's progress, further education needs, or readiness to enter a college.[1]

After compulsory education was introduced in the late 1800s, student enrollments rapidly increased. As a result, descriptive evaluations eventually gave way to grades devised to compare students and thus to sort the faster learners from the slower ones. As interest in attending college and the ability to pay for higher education increased, tests began to be used to determine who could attend.[2]

Testing and grades set up many students for failure in a system that forced all children to attend school, entering at progressively younger ages, and to learn more at an increasingly faster pace, regardless of their ability, need, or interests.[3]

Historically, teachers' evaluations were based on the performance of their students. The teacher's assurance of more students and even survival in the profession was directly linked to his or her pupils' eventual success in the real world. With the adoption of a grading system in the 19th century, students'

success was judged by the teacher, rather than on real-life success after graduation. Classroom learning became driven by teacher-centered evaluation rather than preparation for life. The teacher was no longer held accountable for student failure because he or she could simply blame the student for low grades.[4]

Standardized tests followed the entrance of the grading system. In 1905, Alfred Binet devised a testing instrument to identify children who needed special training. A decade later in the United States, Arthur Otis developed instruments that could be administered to a group rather than just one individual. His work led to the development of the Army Alpha, the first general ability test given to a group, used for sorting army personnel during World War I. The perceived success of this instrument led to standardized testing for the nation's schools. Of the hundreds of tests that followed in the next few decades, only a few survived, such as the Stanford Achievement Tests and the Iowa Every-Pupil Tests of Basic Skills.[5]

Today, standardized testing has grown into a major industry. This article will examine standardized tests, known as norm-referenced achievement tests, as well as criterion-referenced tests, and provide recommendations for responsible classroom assessment.

Picture Removed

Norm-Referenced Tests (NRT)

Norm-referenced tests (NRT) assume that "the trait or ability to be measured is assumed to be present in varying degrees in different individuals." Such tests seek "to order those individuals on a continuum ranging from highest to lowest in terms of the amount of that ability the learner possesses."[6]

Standardized tests suffer from several flaws:

1. Assumptions About Normal Distribution of Scores

Based on the above philosophy, standardized tests purposely set out to produce a wide range of evenly distributed scores. Test makers assume that any given cluster of people contains a predictable percentage of individuals who are "gifted" or "above average," as well as those who are "below average" or "mentally deficient." Test makers designate a certain percentage of correct scores as the mean (average), thereby defining who is above or below average.

The distance from the mean is measured in equal units called "standard deviations." The distribution of scores is often referred to as the bell curve, which places the high and low scores at the extremities and the average scores in the middle. Since half the students are considered to be "above average" and half "below average," test items are selected with the goal that half of the stu-

dents will incorrectly answer or omit the item. This built-in exaggeration of student differences counters the best that we know in psychology and common sense.[7] Furthermore, with a small group of students (such as in most Adventist schools), the scores are not likely to be distributed in the expected bell curve arrangement.

Test publishers begin the standardization process of a new test by trying it out on a group of students.

If a question is answered correctly mainly by the "better" examinees, it is a good question. If it is answered correctly mainly by the "poorer" ones, it is a bad question . . . [Q]uestions are rejected, others are . . . rewritten, . . . and the numerical balance between questions of various grades of objectively determined difficulty is adjusted to the percentage of candidates that the test is designed to reject.[8]

After being tested, students are classified as either above or below average, gifted or mentally retarded.

2. Assumptions About Student Capabilities

In effect, standardized or norm-referenced tests assume that, under normal circumstances, many students have not and cannot master the information that testing firms think should be tested. It has been suggested that Darwinism is partly to blame for this limited perception of human potential.[9]

Experience and research suggest that students can and will learn when given the appropriate experiences and enough time to master the content and skills required of them.

3. Problems of Authenticity

Test designers' concern for predictability typically makes them sacrifice validity for reliability. Tests thus measure "mastery" of isolated data that theoretically represents students' grasp of vast amounts of information, much of which has little or no bearing on their lives or relevancy in the real world.[10]

What are the effects of such tests on the intellectual quality of what is taught and learned?

First, a multiple-choice item does not ask students to produce a solution to a problem or an answer to a question. Rather, it asks them to recognize a solution or an answer. Recognizing and producing are fundamentally different abilities . . . . Do we really want to tell students that being able to recognize the correct answer to a math question will sufficiently prepare them to use math in their everyday affairs?

Second, and even more destructive, is the implicit message that all issues worth discussing and examining can be reduced to selecting among four or five alternatives.[11]

Traditionally, conventional education and generic, mass testing have dealt mainly with mastery of theoretical knowledge—something that occurs in isolation from the real world. The assumption here is that if something is known, it can be applied in context.

Figure 1. Concerns About Testing by North American Division K-12 Educators*

Percentage of respondents answering "Definitely Agree" or "Somewhat Agree" with these statements:

Items From *Profile '93*	Elementary Teachers n=451	Academy Teachers n=203	Conference Personnel n=93	NAD Curriculum Committee n=25
Standardized tests do not measure the goals and objectives of Adventist education	63.1%	50.0%	52.2%	56.0%
Standardized tests do not sample the full spectrum of students' abilities	59.1	44.6	66.3	84.0
Not satisfied with the testing time	40.8	19.7	20.6	16.0
Students become "test weary"	36.7	33.9	NA	NA
Use of test results do not really benefit students	30.2	43.2	27.5	64.0
Standardized tests do not match what I teach	27.8	29.4	NA	NA
Results are not important to parents/constituents	15.2	16.7	4.4	4.0

**Paul S. Brantley, author,* Profile '93 *(Silver Spring, Md.: Office of Education, North American Division of Seventh-day Adventists, 1994).*

This causes educators to "teach to the test" and emphasize rote memorization of fragmentary bits of information. Such instruction does not encourage comprehension, application, or synthesis.

When tests determine what is taught, the curriculum narrows. Children are given fewer opportunities to develop their strengths and must spend more time in drills for multiple-choice exams. Occasions for children's initiative are reduced. They are encouraged to be more passive, more obedient to authority, and less enthusiastic about asking and answering open-ended questions. The "right answer" becomes the goal, and children are left practicing isolated test-taking skills.[12]

In reality, intellectual performance takes place in unique, real-world situations.

Performance is never the sum of drills; problems are not exercises; mastery is not achieved by the unthinking use of algorithms. We cannot be said to "understand" something, in other words, unless we can employ it wisely, fluently, flexibly, and aptly in particular and diverse contexts.[13]

Education is an inherently personal enterprise, and most societies are complex and pluralistic. Therefore, education must be customized to the distinctive needs of each individual student and the diverse cultures in society.

Ultimately, what we really want to know is how competent students are at "thinking, planning, inventing, reassuring, leading, working with others, or developing their own understanding of how society works and taking steps needed to influence it."[14]

Tests are needed that will evaluate the entire range of complex and diverse human abilities.[15] For example, intelligence should be tested in real-life situations where a person is faced with a problem and must generate the questions himself or herself.

Furthermore, standardized tests are not especially effective in measuring some things that Christian schools consider vital to a wholistic education, such as critical thinking and application skills, much less moral responsibility. (See Figure 1 on page 20.)

4. Discrimination

An ongoing concern about standardized tests is that they discriminate against certain groups:

- Persons whose thinking is comprehensive and divergent, who see more than one possible answer, and thus select different answers—often better ones—than they were supposed to.[16] Standardized tests thus favor those who are accustomed to quick, superficial, and convergent thinking.
- Right-brain thinkers, whose mental processes favor analysis (viewing the parts) over synthesis (viewing the whole).
- Persons whose skills lie outside the linguistic and logical-mathematical areas.[17] Tests traditionally measure and reward narrow, intellectual performance skills. We must find ways to teach and assess esthetics, physical skills, social skills, and ability to work with others.
- Immigrants or others who have a limited familiarity with the English language or American culture.
- Persons with learning disabilities.
- Those who suffer panic attacks when confronted by tests or who have a problem completing the questions in the allotted time.
- Females, minorities, and low-income students. A New York federal judge ruled that the Scholastic Aptitude Test is biased against women. While girls score lower on this college entrance exam than boys, particularly in mathematics, in the classroom girls do as well as boys or better.[18] The National Center for Fair and Open Testing says that the SAT is biased against female, minority, and low-income students, and should be eliminated from all college admissions processes.[19]

Tests make no allowance for cultural or geographical differences, either. A group of inner-city children took a standardized test that asked, "Which lives closer to you—a giraffe or a cow?" Because all the children had seen a giraffe at the city zoo but had never seen a cow, they marked "giraffe." The test designers assumed that the chil-

Picture Removed

Since half the students are considered to be "above average" and half "below average," test items are selected with the goal that half of the students will incorrectly answer or omit the item.

dren would think of the natural habitats of cows and giraffes, so all the children got the question wrong.

School entrance age and pupil maturity, as well as different levels of performance in various subjects, can also affect test performance.[20]

5. Test Clarity and Quality

By design, multiple choice tests are ambiguous, so that right answers are not too easily distinguished from wrong ones. Nevertheless, as long as only one answer is clearly correct, this does not pose a problem. But knowing where to draw the line is an imperfect science. Consider the following item, taken from an actual test:[21]

A requirement for a democratic government is that:

a. the people must have a voice in government

b. government leaders must do what they think is best for people

Of these two items, which is correct? When the examinee is asked to select the "best" answer in similarly constructed tests, this can cause considerable difficulty.

Standardized tests have also been criticized because students receive no feedback on their answers, cannot request clarification of confusing questions, and are not allowed to elaborate on their choices.[22] Also, tests make unreasonable demands upon students, including the need to work well under pressure, to solve problems quickly, and to have a high toleration for boredom.[23]

6. Test Preparation and Administration

It is only natural that teachers and administrators who are evaluated on the basis of how well their students do on national test results should be unnerved by the thought of low scores. These fears lock educators into traditional, ineffective educational practices and tempt them to use an array of disreputable approaches to produce high test scores. John Cannell's book, *How Public Educators Cheat on Standardized Tests*, presents the result of a national study of state reports of students' achievement. After obtaining results from more than 3,500 school districts, Cannell's team concluded that 70 percent of American school children, 90 percent of American school districts, and all 50 states were testing above publishers' "national norm" on commercial norm-referenced achievement tests. Educational testing was found to be politicized and to grossly overstate achievement. Cannell's findings were later confirmed by the Federal office of education.[24]

Cannell discovered certain activities associated with dubious achievement testing, including: (a) teaching to the test, (b) aligning the state curriculum with the tests, (c) selecting tests that will yield the highest scores, (d) excessive coaching before and during testing sessions, (e) using loopholes to "legally" prohibit low-achieving students from taking the tests, (f) student cheating, (g) biased and deceptive interpretation, and (h) communication of the scores to the media.[25] These teaching and testing practices demonstrate the pervasive concern that any universally used standard measurement will drive local curriculum and classroom pedagogy, and that schools will rise no higher than the expectations placed upon them by standardized tests.

Use of Test Scores

Concerns about use of test scores are varied: They drive public policies and decisions about spending on education, affect real-estate values, and lead to comparisons between districts, states, and countries. Standardized tests serve

Guidelines for Administrators and Teachers Involved in Standardized Testing

Administrators (and Staff)

1. Develop curriculum by using criteria other than national achievement tests.
2. Develop well-articulated school standards of academic excellence, which you guarantee to your constituents, and to which you hold yourself and your students accountable.
3. Encourage teachers to teach according to school standards and student needs.
4. Exercise caution in releasing conference or union test results that encourage undue emphasis on achievement tests and inter-school competitiveness.
5. Make provisions to bring staff together to discuss central questions of assessment design and standards for interpretation of data.[26] Then provide in-service training, coaching, and sufficient time for teachers to understand and implement the new procedures successfully.[27]
6. Provide workshops, in-service training, and time for all the constituencies to discuss and understand new assessment methods and their implications.

Teachers

1. Test sparingly.
2. Be cautious in your interpretation and communication of test results to the parents. Such results never apply to an individual student. Refrain from undue criticism of lower scores or praise of higher scores.
3. Do not label students on the basis of their test results.
4. Inform those concerned about test scores that determination and perseverance are two of the most important determinants to student success.
5. Promote peer tutoring and cooperative learning that strengthen relationships and suppress the competitive spirit fostered by tests.
6. Develop quality criterion-based tests and alternative assessments such as portfolios that meet your students' individual needs, allow for creative and diverse methods of evaluating and reporting, and make winners out of all your students.
7. Remind yourself often that each student is a child of God, made in His image, with enormous potential for development and service to humankind.

as gatekeepers for entry into college and professions, determine who gets scholarships and grants, and have been used to stereotype certain races and cultural groups.

As the Scriptures warn us (2 Corinthians 10:12, NRSV), "We do not dare to classify or compare ourselves with some of those who commend themselves. But when they measure themselves by one another, and compare themselves with one another, they do not show good sense." Christian educators must sense the moral and academic imperative to seek more humane, fair, and discerning forms of assessment for our youth. (Other articles in this issue suggest some alternative forms of assessment.)

North American Adventist K-12 educators surveyed in *Profile '93* expressed a variety of concerns about standardized testing. (See Figure 1.) The most significant of these were that (1) the tests did not measure the goals and objectives of Adventist education, and (2) they failed to sample the full spectrum of students' abilities. We need evaluation measures, combined with a relevant, wholistic, personalized education that assumes that all children can learn and master those skills necessary for a meaningful life on this earth and an eternal reward in the life hereafter.

Jerry Kea has served as a classroom teacher, administrator, and college professor, and has been associated with home schooling as well as Adventist and public education. He currently is promoting H.O.P.E., a public benefit corporation that will offer support services to schools of choice throughout the United States, and writes from Berrien Springs, Michigan.

REFERENCES

1. Howard Kirschenbaum, Sidney B. Simon, and Rodney W. Napier, *Wad-ja-get? The Grading Game in American Education* (New York: Hart Publishing Company, Inc., 1971).
2. Ibid.
3. Raymond S. Moore and Dorothy N. Moore, *Better Late Than Early* (New York: Reader's Digest Press, 1975).
4. Kirschenbaum, et al.
5. James D. Linden and Kathryn W. Linden, *Tests on Trial* (Boston: Houghton Mifflin, 1968).
6. D. A. Gilman, *Alternatives to Tests, Marks, and Class Rank* (Terre Haute, Ind.: Curriculum Research and Development Center, School of Education, Indiana State University, 1977), p. 11.
7. Ibid.
8. Banesh Hoffmann, *The Tyranny of Testing* (New York: Crowell-Collier Press, 1962), p. 55.
9. Grant P. Wiggins, *Assessing Student Performance: Exploring the Purpose and Limits of Testing* (San Francisco, Jossey-Bass Publishers, 1993).
10. Ibid., p. 43.
11. Judah L. Schwartz, "The Intellectual Costs of Secrecy in Mathematics Assessment," in Vito Perrone, ed., *Expanding Student Assessment* (Alexandria, Va.: Association for Supervision and Curriculum Development, 1991), p. 137.
12. Kathe Jervis, "Closed Gates in a New York City School," in Perrone, p. 4.
13. Wiggins, p. 207.
14. J. Raven, *The Tragic Illusion* (Unionville, N.Y.: Trillium Press, 1991), p. 36.
15. Hoffmann, p. 30.
16. Ibid.
17. Howard Gardner, *Frames of Mind* (New York: Basic Books, 1983).
18. See http://www.daily.umn.edu/gopher-archives/1989/03/02/Change-S...
19. See http://www.dailynorthwestern.com/daily/issues/1997/11/20/cam...
20. Moore and Moore.
21. Wiggins.
22. Ibid.
23. Gene D. Shepherd and William B. Ragan, *Modern Elementary Curriculum* (New York: Holt, Rinehart, and Winston, 1977).
24. J. Cannell, *How Public Educators Cheat on Standardized Achievement Tests* (Albuquerque: Friends for Education, 1989), p. ix.
25. Ibid.
26. Edward Chittenden, "Authentic Assessment, Evaluation, and Documentation of Student Performance," in Perrone, p. 23.
27. George E. Hein, "Active Assessment for Active Science," in Perrone, p. 128.

Picture
Removed