

The Evolution of Large-Scale Educational Assessment in the United States

Mark D. Reckase

Michigan State University



My Own Misunderstanding of History

- My Ph.D. is in psychology and I knew little about educational assessment until getting my first teaching position.
 - Even though my advisor was an author of the Stanford Achievement Test and Metropolitan Achievement Test
 - At that time tests were produced by authors and sold by publishers like books.
- My first teaching job assigned me to teach testing to teacher candidates, so I learned about educational assessment.
- I thought that what I learned at that time was the way it always was and always would be.



Achievement Testing when I Started My Career

- Numerous commercial publishers distributed tests like textbooks.
- They marketed the tests plus reporting services.
- Each school district selected the commercial test that they thought best matched their needs.
 - My first consulting job was helping a school district select the commercial testing package that they would use.
 - It was a very unhappy experience
 - Tried to determine how they would use the test.
 - They did not like my recommendations.



Motivation for this Presentation

- Share a big-picture view of educational testing.
- Requested to do an encyclopedia entry on the history of large-scale achievement testing since 1980.
 - I thought this would be easy and fast.
 - But I decided to look at the full history of educational testing starting with colonial times.



Testing at the Time of the Founding of Our Country

- “There was a midday intermission from eleven to one, except on Monday, when the master shall call his scholars together between twelve and one of the clock to examine them what they have learned, at which time also he shall take notice of any misdemeanor or outrage that any of his scholars shall have committed on the sabbath, to the end that at some convenient time due admonition and correction shall be administered.”
- Dorchester, Massachusetts, 1645



First Large-Scale Achievement Testing: 200 Years Later

- Prior to 1845, in Boston, a committee would visit each school for a day to evaluate how well the school was functioning.
 - Invariably, they would give a positive report.
- At the encouragement of Horace Mann, Secretary of the State Board of Education, Massachusetts, the School Committee of the City of Boston decided to give the students a test.
- Horace Mann was also the editor of *The Common School Journal* and he published the results.
- The general conclusion was that the students were not doing as well as the committees had reported.



Example Mathematics Items

- A stationer sold quills at 10s 6d per thousand, by which he cleared $\frac{1}{2}$ of the price – but the quills growing scarce, he raised the price to 12s per thousand. What percent would he clear by the latter price?
- How much is $\frac{1}{2}$ of $\frac{1}{3}$ of 9 hours and 18 minutes?



Example Natural Philosophy (Science) Item

- Which could you stop most easily, a railroad car weighing a ton going at the rate of 10 miles an hour, or a car weighing 100 tons, creeping along at the rate of $\frac{1}{4}$ of a mile an hour?
- In the diurnal revolution of the earth, who are moved with greatest velocity, the inhabitants of Mexico or of Boston?



Testing at the Turn of the 20th Century

- There was great interest in making education more scientific.
- In Kansas in 1902, Cornman was trying to determine if the amount of instructional time spent on spelling was related to the level of achievement.
 - Discovered that instruction on spelling could be reduced without a loss in spelling achievement.
- Thorndike in 1903 indicated that knowledge about student capabilities was needed to properly plan instruction.



Influence of the Army Alpha and Beta

- Education in early 1900s was uneven.
- The US entered World War I in 1917 and a committee of psychologists was called together to develop a way to screen draftees according to how quickly they could learn about military activities.
- Created the Army Alpha for those literate in English and the Army Beta for those who were illiterate.
 - The tests were very successful for placing persons into military occupations.
 - The tests were administered to over a million persons and could be scored very quickly. It was the beginning of “objective” scoring.
- In 1920, after the war was over, the committee published the tests and the results of their use.



Influence of the Army Alpha and Beta

- During the 1920s and 1930s, authors created school versions of these tests and book publishers sold them to schools.
- The Stanford Achievement Test (Terman at Stanford University) was probably the first of these in 1922.
- Commercial intelligence tests were also produced – The Otis Quick-Scoring Intelligence Test.
- Each school would decide if they wanted to use any of these tests and, if so, which one.



Evolution of Standardized Testing from the 1930s to 1970s.

- During that time, six major publishers were actively marketing standardized tests to schools.
- The Scholastic Aptitude Test (SAT) was first used in 1926 and the ACT in 1959.
- The National Assessment of Educational Progress (NAEP) came into existence in 1969.
 - Often labeled a survey rather than a test.



Events that Influenced the Form of Educational Testing in the 1980s.

- Sputnik was launched in 1957. The event stimulated concerns about the educational system in the U.S.
- California was particularly concerned. In 1960, a Commission approved that standardized tests should be administered in all the schools in the state.
 - Tests were first administered in 1963.
 - Schools could choose what tests to use.
- About the same time, the Iowa Testing Program was started.
 - It was used by most schools in the state of Iowa.
 - It was not required.
- Florida started state testing in 1971.



Problems with the Tests Used in the Schools

- The “Lake Wobegon effect” raised questions about the credibility of commercial standardized tests.
 - Cannell (1987) found that all states reported that their students were above average.
 - Phillips and Finn (1988) coined the term to describe the results in a reaction to the results.
 - "Well, that's the news from Lake Wobegon, where all the women are strong, all the men are good-looking, and all the children are above average."
- A 14-year decline in SAT scores (1977) was noticed and was of great concern.
 - Many analyses were done to identify the cause of the decline.
- The result was a lack of confidence in the information provided by standardized tests.



Responses to the Credibility Issues

- Secretary of Education, Terrel Bell, started collecting data from the states and reporting them on in a large format document called the “Wall Chart” (1984).
 - It gave rankings of states on many variables.
 - The Wall Chart was criticized on many technical grounds because the different states used different tests and the samples of students taking the SAT and ACT were very different.
- The governor of Tennessee, Lamar Alexander, chaired a committee that recommended that NAEP report results at the state level (1988) to give a credible comparison of the state educational systems.
 - He also began using value-added models to evaluate teacher performance.
- In 1990, NAEP added standards of performance called achievement levels – in particular Basic, Proficient, and Advanced.



Consequences of the Attention to Large-Scale Testing

- State NAEP results were sometimes inconsistent with the results reported on state assessments.
- The No Child Left Behind Act of 2002 required states to have a state testing program that included proficiency standards to get federal funding.
 - Required states to define “Adequate Yearly Progress.”
 - Testing program must be approved to get funding.



Educational Assessment Is Now Very Complex

- Educational assessment is now dominated by state assessment programs, but these sometimes have complex component parts.
 - ACT or SAT
 - Smarter Balanced or PARC
- National and international programs have a lot of visibility
 - NAEP
 - PISA
 - TIMSS
- Tests from publishers and independent agencies are still in the mix.
 - Iowa Tests
 - NWEA MAP



Current State of Affairs

- State testing programs are required to have peer review that requires extensive information from the state.
 - Alignment of test to state curriculum standards has become a major part of peer review.
 - The use of SAT, ACT, and the tests from the consortia complicate alignment evaluations.
- The role of commercial organizations has changed.
 - Originally like book publishers
 - Changed to producers of content – testing companies
 - Now changing to technology companies with the delivery of tests on computer.
- There is increased emphasis on international comparisons with TIMSS, PISA, etc.



Directions for the Future

- More emphasis on formative assessment
- Desire for more subscores and diagnostic information.
 - Sometimes with shorter tests and less testing time.
- Use of technology to enhance connections of tests to instructional goals.
 - Sometimes with computerized adaptive tests.
- Emphasis on collection of data to develop a validity argument for the inferences to be made from scores.
 - Requires clear presentation of supported inferences.



Conclusions

- My only confident conclusion is that technology will have a bigger impact on testing.
- Everything else is uncertain because of the continuing trends for legislators to put requirements on testing programs.
 - Blame California, Secretary Bell and Governor Alexander.
 - No Child Left Behind



Final Thoughts

- Stability of system can result in improvement in quality.
- Continuing tension between Federal involvement in education and the desire for local control.
 - Educators in other countries do not understand the educational system in the U.S.
 - Few people in the U.S. understand the educational system.

