



A partnership among Keystone College, King's College, Luzerne County Community College, Marywood University, Misericordia University, Penn State Wilkes-Barre, The Commonwealth Medical College, University of Scranton & Wilkes University

Standardized Testing

Table of Contents

INTRODUCTION 3

SELECTED ARGUMENTS FOR STANDARDIZED TESTS..... 3

SELECTED ARGUMENTS AGAINST STANDARDIZED TESTS..... 5

ALTERNATIVES TO STANDARDIZED TESTS 7

CASE STUDIES ON THE MEASUREMENT OF STUDENT PERFORMANCE 9

POLICY RECOMMENDATIONS 10

POLICY STATEMENTS 11

CONCLUSION 11

Introduction

The purpose of standardized tests is to compare the abilities and skills of students from different schools and backgrounds.¹ They are designed to be a neutral and fair examination of test participants. Standardized tests are used to measure aptitude or knowledge across a multitude of schools or contexts.² This means that an entity outside of the school or education system is responsible for determining how well students and schools are doing relative to other students and schools. The importance and consequences of standardized tests have been increasing over the last ten years. The focus of this paper - the No Child Left Behind (NCLB) law is largely responsible for this focus on standardized testing.

NCLB was signed into law in 2001 by President George W. Bush. The law requires each state to develop tests to measure basic skills, such as reading and math comprehension. Such tests are administered at specific grade intervals. The primary purpose of NCLB is to ensure that students in every public school achieve important learning goals, while being educated in safe classrooms by well-prepared teachers.³ As such, there has been much debate surrounding the most effective way to increase test scores. There has also been significant debate concerning the validity of the state-issued standardized tests.

NCLB requires that states administer tests to all public school students. The states set proficiency standards, called adequate yearly progress, that progressively increase the percentage of students in a district that must meet the proficiency standard.⁴ If a school district does not meet these proficiency levels, the law mandates that requirements be met and corrective actions applied.⁵ Usually, the law mandates incentives like increased or decreased funding based on test results. This means that the government could cut funding to a school if it does not perform well on the test. This is very controversial, as it calls into question the overall purpose and effectiveness of NCLB and standardized tests.

In recent years, there has been a push to reform NCLB. President Obama has proposed more in depth testing. He has also granted several states waivers from the law, allowing for more control over assessments. The Forum on Educational Accountability also identified several NCLB reforms, which fall under three categories: “Educationally Helpful Assessments,” “Rational Expectations for Improvement” and “Support Instead of Punishment.”

Most standardized tests consist of multiple choice and short essay answers. There are many arguments supporting the effectiveness and need for standardized tests in the school system. However, there are just as many arguments supporting why standardized tests are ineffective for students, schools, and education itself. Fortunately, there are alternatives that can take the place or aid the validity of standardized testing. Most of these alternatives call for more interaction with the assessors and a sharper focus on student improvement rather than student comparison.

Selected Arguments for Standardized Tests

Standardized tests are designed to compare. They are designed to see how a student, school, technique or region compares to others in the same category. Although there is debate as to

what these tests actually compare (intelligence, knowledge, test taking abilities), it is important to note that such tests are consistent in their measuring and do offer some valid comparisons.

Standardized tests are reliable and objective measures of student achievement. Without them, policy makers would have to rely on tests scored by individual schools and teachers, who have a vested interest in producing favorable results.⁶ In other words, the government cannot rely just on student grades to determine a student's performance compared with every other student in an area (i.e., city, county, state, region, country).

For example, if school offers two American History courses and each class is taught by a different teacher, the teachers likely have different tests, education techniques, and possibly a different grading scale. One could conclude that one American History course should be slightly more difficult than the other based on all the variables. If Student A is in the first class and Student B is the second class receive the same grade at the end of the year, but Student A's class is more difficult than Student B's class, is it right to assume that they know the same amount of material? What if Student B received a higher grade than Student A? Would it be safe to say that Student B was smarter than Student A in American History even though Student A was in the more difficult class? Clearly, it is easy to see the complications of just using grades to compare students. Note that the above example compares two students in the same school. This problem is compounded when you examine any two students in the country. Since standardized tests, by definition, are standard, there is no confusion about how two students compare to each other.

It is also important to note that standardized tests provide useful information at a very low cost, and consume little class time. Standardized tests cost less than 0.1% of K-12 education spending, totaling \$5.81 per student per year.⁷ Students are expected to take and do well in standardized tests throughout their whole academic career, and are the most popular way to evaluate students at a snapshot in time. The U.S. Department of Education stated, "Although testing may be stressful for some students, testing is a normal and expected way of assessing what students have learned."⁸ Students are expected to be able to communicate what they have learned in a test. As such, this is what they are doing when they take a standardized test; it simply focuses on more generalized material.

There are three levels of standardized testing that a student could face even if high school and grade school assessment testing were eliminated. The first level is the college/graduate entry test, including, for example SATs and GREs. These are tests used to compare students with other students when a student is looking to further his/her academic career. Student test results are compared to other students seeking the same level of education. Colleges and universities use these scores as an important admissions factor.

The second level of standardized testing is the professional entry test, intended for those who wish to go to a professional school. Examples include MCATS (medical school) and LSATS (law school). Although these tests can be topic specific, they are still standardized. Individual results are compared against every other student who takes the test. Professional schools use these scores as an important admissions factor.

The third level of standardized testing is the professional certification test. Examples include the PRAXIS and BAR exam. Professional certification tests are designed to ensure that students know enough information about a certain profession so they can legally perform the job they are seeking. Physicians, lawyers, real-estate brokers and pilots all take high-stakes standardized tests to ensure they have the necessary knowledge for their professions. If standardized tests were an unreliable source of data, their use would not be so widespread.⁹

As a side note, in order to become a naturalized citizen in this country, foreign citizens must pass a standardized test.

Selected Arguments against Standardized Tests

Education encompasses a wide range of ideas, concepts and information. It is nearly impossible to test a student on every aspect of his/her education. This means that a lot of subjects are excluded from standardized testing.

In order to receive a diploma, students must take a variety of classes in different subjects, including history, social studies, geography, arts and physical education. While these subjects are not included on standardized assessment tests, students are required to pass them in order to graduate. Mathematics and reading are the only two subjects tested in every state. Other subjects, like science and writing, are tested in some, but not all, states. Countless other subjects are not tested through standardized assessments.

Educator Bill Ayers said, "Standardized tests can't measure initiative, creativity, imagination, conceptual thinking, curiosity, effort, irony, judgment, commitment, nuance, good will, ethical reflection, or a host of other valuable dispositions and attributes. What they can measure and count are isolated skills, specific facts and functions, the least interesting, and least significant aspects of learning¹⁰."

Imagine a student who excels in writing historical fiction. Such student is able to show vast knowledge in history to remain accurate in his/her writings. Also, this student shows excellence in the area of creativity and imagination - to continually write stories at a high level. In fact, the student looks to further his/her career in this field. With standardized tests in most states, the student will not even be assessed in what makes him/her a good student. If the student's math skills are poor, then he/she may be considered a below average student.

Who is responsible for low or high student test scores? The assessment system says that test scores are teacher and school responsibilities. Although education comes from the teachers, parents, and students themselves, the consequences of this high stakes testing falls mainly on the shoulders of teachers and schools. If a school fails to meet the level of proficiency that is required under NCLB, the school could face major consequences, including the loss of funding and being mandated to teach for longer hours and stay after school to provide free tutoring.¹¹ If a school continues to fail to meet the level of proficiency required by NCLB, the school may be forced to replace its teaching staff, school administration, or, in the most extreme case, it may be forced to close.¹²

It is universally known that parents are the best teachers and help children learn throughout their lives.¹³ Under NCLB, parents or guardians are not required to attend any parent-teacher conferences to discuss their child's progress. And while parents are not punished when their child does poorly, they are neither rewarded when their child does well. In fact, parents play no role in the standardized test system. This is unusual since few people would argue that a large majority of what students learn takes place in the home or through homework.

Likewise, students who actually taken standardized tests face few consequences for poor performance. In fact, only twenty three states have graduation requirements in place, which require a student to pass a standardized test. This means that a student who fails a standardized test in which teachers' and school administrators' jobs are at stake, may still graduate without knowing the material that the laws requires such student to know. A teacher can be fired before the student is held back for failing such test.

Education is a three legged stool.¹⁴ In addition to schools and teachers being responsible for education of children in the U.S., parents are as well. As such, it makes no sense that all of the responsibility of such education is placed on standardized testing.

Another issue with standardized testing is the amount of time it takes for schools to receive assessor feedback. In Pennsylvania, test scores are not released to schools until September.¹⁵ By this time, students who have taken the tests have already scheduled their classes and been participating for almost a month in the classes they selected. Meanwhile, those students the grade level below are now learning the same curriculum in math and reading as the student before them. One can easily see an issue in this situation if the first set of students did not perform well on a certain part of the standardized test.

To help illustrate this point, consider a set of students (Students A) take a standardized test. Students A do not get their test results back until they have scheduled and started taking the next year's classes. The next set of students (Students B) starts their school year with the same curriculum as Students A in the prior year. The school gets test results back and Students A did not meet the level of proficiency required in mathematics. It is now very difficult for Students A to learn what they struggled with in the previous year since they are already enrolled in a new course. It is also difficult for teachers to change the curriculum once it is underway. While teachers can change little things in the curriculum, major changes must wait until the following school year. Thus, Students B will learn from the "bad" curriculum.

The following school year brings in Students C. The curriculum has changed and this set of students is learning the areas of mathematics that Students A struggled with. The school gets the results back and, because they were unable to make major changes to the curriculum, Students B failed to meet their proficiency level, too. The school hopes that the changes will increase test scores but it will not find it such changes have done so until the following school year with Students D. This means that if a school fails to meet a proficiency level in one subject, it would take at least four years to know whether or not the school is able to fix the problems in that subject because of the length of time between the test and the results. It is also important

to remember that during these four years, another subject might fall under the proficiency level and require more time to address.

The standardized testing system creates the idea of winners and losers. If a student passes the test, he/she is a winner. If a student fails the test, he/she is a loser. It doesn't matter how many points the student scored over or under the proficiency level, it simply matters on which side of the pass/fail line he/she falls. This creates an attention gap among students. According to Daniel Koretz's study, *What Do Standardized Tests Really Tell Us About Students and Schools?*, "It creates really bad incentives. And we're now beginning to slowly accumulate research confirming that when you reward teachers for getting kids across one line, they focus on the kids who are near the line to the detriment of others."¹⁶ This means that students who are above the proficiency level may be ignored. Likewise, those who are below the proficiency level may not receive much attention because schools realize that there are not enough resources to bring these students up to the proficiency level. This leaves the most resources and attention going to students who are close to the proficiency line.

The study also stated that teachers and schools have "figured out what's now called the bubble kids, who the bubble kids are, the kids right on the cusp, and moves a bunch of them from just below that cut score to just above will look like they're making huge progress."¹⁷

Alternatives to Standardized Tests

With so much debate over standardized testing, many alternatives have been proposed. One such alternative is the use of portfolios - a collection of evidence, prepared by the student and evaluated by the faculty member, to demonstrate mastery, comprehension, application, and synthesis of a given set of concepts. To create a high quality portfolio, students must organize, synthesize, and clearly describe their achievements and effectively communicate what they have learned.¹⁸

The government can still assess students using portfolios by establishing criteria for the materials that all students must include. The government could mandate items such as an essay on literary text, a geometry math proof, and biology lab report, music interpretation and whatever else it feels is appropriate for the grade level being assessed. Each student can complete his/her portfolio and submit it for assessment.

There are several benefits to this alternative. Portfolios put the responsibility of demonstrating knowledge and integration across concepts on students.¹⁹ They also provide a structure for long-duration assignments.²⁰ Portfolios encourage student creativity and allow them to emphasize the aspects of a concept most relevant to them in meaningful ways.²¹ Finally, portfolios engender self-reflection and self-assessment.²² This means that they help to solve the "can't test everything" problem of standardized tests. Government officials are able to choose exactly what they feel is important to assess.

Portfolios, however, also have their problems. They take longer to score than machine graded multiple-choice exams.²³ This may cause a greater time lag than previously discussed regarding standardized tests. Portfolios involve student work outside of class.²⁴ Although this would put responsibility on parents, it might put too much responsibility on parents. Portfolios do not easily demonstrate students' knowledge-recall abilities.²⁵ Finally, students who have been successful at memorizing their way to an "A" initially find portfolios intimidating.²⁶ So while the portfolio solves the issue of standardized testing not being able to test everything, they present both time and responsibility concerns.

Another alternative to standardized testing is the interview. An interview is a face-to-face assessment that consists of a series of well-chosen questions (and often a set of tasks or problems), which are designed to elicit a portrait of a student's understanding about a scientific concept or set of related concepts.²⁷ The interview may be videotaped or audio taped for later analysis.

Government officials can use this alternative to standardized testing by establishing criteria for the subjects and concepts it wants each student to know at each grade level. The government might even choose to mandate a strict set of questions to be asked of each student. Each student might go through three sets of interviews – one of which is with a teacher from the school the student attends. This offers immediate feedback to the school and student. The second interview might be given by a teacher from another school district (preferably a close by district). This may eliminate bias that may occur with a teacher that a student knows and still allows for quick feedback to the school and student. The third interview might be given by a government assessor. This could offer the national comparison sought, while completely eliminating any bias.

There are many benefits to the interview assessments. Structured interviews enable instructors to examine the degree of understanding a learner has on a scientific concept or closely related group of concepts.²⁸ Interviews offer instructors a vehicle for focusing on how instruction is interpreted and internalized by the learner.²⁹ Faculty can use structured interviews as a powerful type of formative assessment to improve courses and teaching methods.³⁰ Collaborative analysis of interviews allows faculty groups to develop shared understandings of typical difficulties students have with key concepts.³¹ Not only do interviews help eliminate the time lag issue, but they also help solve the "can't test everything" problem of standardized tests.

Drawbacks to the interview assessment is that they are designed to elicit how a student understands a scientific concept. As such, they should be used in addition to, not instead of, other forms of evaluation.³² Interviews are quite time-consuming. Faculty members should interview a broad sample of students in a class in order determine how they are reacting to and understanding the concepts presented.³³ The usefulness of the interview technique is largely determined by the nature and quality of the probes and follow-up questions.³⁴ Thus, a substantial amount of planning may be required to design an informative interview.³⁵

Like portfolios, interviews allow for assessors to assess a wide range of topics, subjects and concepts. Interviews also help solve the time issue of standardized tests by encouraging somewhat immediate feedback. However, the most difficult part of interviews is the student comparison. This is why interviews should be used with another form of assessment instead of a replacement.

Case Studies on the Measurement of Student Performance

This section outlines case studies that demonstrate why student performance should be based on more than one measurement.

Case Study 1: High-Stakes Testing and Mathematics Performance of Fourth Graders in North Cyprus

The authors attempted to determine the effects of a high-stakes standardized testing on mathematical performance. They developed a multiple-choice mathematics performance test for 1,000 fourth grade students in 28 schools. Overall results show that the fourth graders' performances in routine mathematics items were better than their performances in non-routine mathematics items. Although the amount of time that students spent preparing for a high-stakes standardized test was an important factor in their mathematics performances, the differences favoring the students who were instructed with a test-driven approach resulted from their answering mostly routine number, operation and story problems. However, relatively small effect sizes indicates that a test-driven instructional approach has little to do with improving higher order skills like solving non-routine story problems. Overall, students who were taught only to perform well on standardized tests showed little ability to perform well on anything outside of testing such as critical thinking math problems. This case study shows not only that teaching to standardized tests does not improve student performance, but is an impediment to critical thinking.

Case Study 2: Dropout Rates after High Stakes Testing in Elementary School

One concern about high stakes student assessment policies is that they could increase dropout rates. Those who support such policy say that adverse effects of grade retention should be more than offset by beneficial effects of rising achievement. Using data from Chicago, this study examines dropout rates after implementation of an eighth-grade promotion standard. The results indicate that retention by the policy did have adverse effects on dropout rates, but the relationship was smaller than seen with traditional teacher-initiated retention and was unrelated to the timing of dropping out. System-wide, slight decreases in dropout rates among the 90% of students who were not retained counterbalanced the higher dropout rates among those retained.

In Chicago, the implementation of an eighth grade promotional standard was not accompanied by a massive rise in overall dropout rates, as feared by policy opponents. However, overall dropout rates also did not decline for several years, despite substantial improvements in student achievement.

Case Study 3: Foundations for Success - Case Studies of How Urban School Systems Improve Student Achievement

This report examines the experiences of three urban school districts that have raised academic performance within their district. The research is based on case studies of these districts and comparisons of their experiences with other districts that have not yet seen similar improvements. The case study districts are used to develop hypotheses about the reasons for achievement improvements. The comparison districts provide a partial test of the hypotheses emerging from the analysis of the case study districts.

The case study districts' approaches to reform had the following in common:

- They focused on student achievement and specific achievement goals, on a set schedule with defined consequences; they aligned curricula with state standards; and they helped translate these standards into instructional practice.
- They created concrete accountability systems that went beyond what the states had established in order to hold district leadership and building-level staff personally responsible for results.
- They supported district-wide strategies at the central office through professional development and support for consistent implementation throughout the district.
- They committed themselves to data-driven decision-making and instruction. They gave early and ongoing assessment data to teachers and principals, as well as trained and supported them as the data were used to diagnose teacher and student weaknesses and make improvements.
- They started their reforms at the elementary grade levels instead of trying to fix everything at once.

Policy Recommendations

Currently, too much emphasis is placed on standardized testing. It has become the mechanism to determine a school's funding, to compare students, and evaluate overall school performance. Given the narrow parameters of standardized tests, this tool should not be the single determining factor in funding, school or teacher evaluation or student performance. This emphasis could cause schools to "teach to the test" as oppose to teaching content. This is a losing proposition for all involved.

Standardized testing should be a component of evaluating progress or for use in comparisons. Emphasis should be on student performance, not comparisons. Student performance should also be a component of teacher evaluation. It should be noted that teacher evaluation is the second policy brief in this series, as the two topics are interdependent.

Policy Statements

The Institute for Public Policy & Economic Development supports student performance being based on more than one form of measurement.

The Institute for Public Policy & Economic Development supports legislation that includes additional methods of performance measurement.

The Institute for Public Policy & Economic Development supports school districts taking a proactive approach to performance measurement

Conclusion

Although the purpose and intention of standardized testing are pure, the true effects of such tests remain uncertain. After careful analysis, there seem to be many challenges with the effectiveness of such tests. Some might argue that many of the problems associated with standardized tests are enough to outweigh the benefits, and, hence, call for the elimination of standardized tests.

Standardized tests are designed to compare. They consistently compare students, school, regions and areas of the nation. They also prepare students for the multitude of standardized tests that they may face while furthering their academic career. If people believe that “practice makes perfect,” then it makes sense that students are exposed to standardized tests early and often.

However, standardized tests are not able to test and measure everything. There are aspects of education, like motivation and creativity, which standardized tests cannot measure. As the system currently operates, too much responsibility is placed on teachers and schools. That is, there is not enough responsibility on students or their parents and guardians. Another issue is the amount of time it takes to receive test results. It could take up to four years to fix a problem in one subject, once a problem is identified and addressed. Standardized tests also promote an attention gap among teachers and students. Students receive help based on close they are to the proficiency line.

Sources

Abbott, J. (2008, January 28). *True Education is a Three-legged Stool*. Retrieved March 2012, from Change Learning: <http://changelearning.ca/media/true-education-three-legged-stool-john-abbott-speaks>

Allensworth, Elaine. *Dropout Rates After High Stakes Testing in Elementary School: A Study of Contradictory Effects of Chicago's Efforts to End Social Promotion*. *Education Evaluation and Policy Analysis*. Winter 2005. Volume 27. Number 4. pp. 341-364.

Camm, A. (2010, December 28). *The Case Against Standardized Testing*. Retrieved March 2012, from Educators' Guide to Innovation: <http://guidetoinnovation.ning.com/profiles/blogs/the-case-against-standardized>

Great Schools. (n.d.). *Consequences for Schools "In Need of Improvement"*. Retrieved March 2012, from Great Schools: <http://www.greatschools.org/definitions/nclb/nclb.html>

Koretz, D. (2009). *What Do Standardized Tests Really Tell Us About Students and Schools?*

No Child Left Behind Reform: What's the Alternative? Retrieved from FairTest.org. <http://www.fairtest.org/no-child-left-behind-reform-whats-alternative>.

Pennsylvania Department of Education. (2012). *2011 PSSA Results Release Timeline*. Commonwealth of Pennsylvania.

ProCon. (n.d.). *Pro Standardized Tests*. Retrieved March 2012, from ProCon: <http://standardizedtests.procon.org/>

Slater, T. F. *Portfolios*.

Smith, M. U., & Southerland, S. A. *Interviews*.

Snipes, Jason, Fred Doolittle and Corinne Herlihy. *Foundations for Success*.

Case Studies of How Urban School Systems Improve Student Achievement. <http://www.mdrc.org/> <http://www.mdrc.org>

Standardized Tests 101. (2009). *Standardized Tests 101*. Retrieved March 2012, from Standardized Tests 101: <http://www.standardizedtests101.com>.

Teacher Test. (2012). *Children Learn A Lot From Parents*. Retrieved March 2012, from Teacher Test: <http://www.bcer.org/children-learn-a-lot-from-parents-are-parents-interested-in-becoming-better-teachers.htm>.

Tut, Mehmet Ail and Osman Cancoy. High-Stakes Testing and Mathematics Performance of Fourth Graders in North Cyprus

Yell, M. L. (2006). The Purpose of No Child Left Behind. In M. L. Yell, *The Law and Special Education* (pp. 180-181). Pearson Education Inc.

Endnotes

-
- ¹ (Standardized Tests 101, 2009)
 - ² (Standardized Tests 101, 2009)
 - ³ (Yell, 2006)
 - ⁴ (Yell, 2006)
 - ⁵ (Yell, 2006)
 - ⁶ (ProCon)
 - ⁷ (ProCon)
 - ⁸ (ProCon)
 - ⁹ (ProCon)
 - ¹⁰ (Camm, 2010)
 - ¹¹ (Great Schools)
 - ¹² (Great Schools)
 - ¹³ (Teacher Test, 2012)
 - ¹⁴ (Abbott, 2008)
 - ¹⁵ (Pennsylvania Department of Education, 2012)
 - ¹⁶ (Koretz, 2009)
 - ¹⁷ (Koretz, 2009)
 - ¹⁸ (Slater)
 - ¹⁹ (Slater)
 - ²⁰ (Slater)
 - ²¹ (Slater)
 - ²² (Slater)
 - ²³ (Slater)
 - ²⁴ (Slater)
 - ²⁵ (Slater)
 - ²⁶ (Slater)
 - ²⁷ (Smith & Southerland)
 - ²⁸ (Smith & Southerland)
 - ²⁹ (Smith & Southerland)
 - ³⁰ (Smith & Southerland)
 - ³¹ (Smith & Southerland)
 - ³² (Smith & Southerland)
 - ³³ (Smith & Southerland)
 - ³⁴ (Smith & Southerland)
 - ³⁵ (Smith & Southerland)