Educational Implications of High Stakes Testing

Nicole S. Caldwell
Regis University

Follow this and additional works at: https://epublications.regis.edu/theses
Part of the Education Commons

Recommended Citation
https://epublications.regis.edu/theses/1

This Thesis - Open Access is brought to you for free and open access by ePublications at Regis University. It has been accepted for inclusion in All Regis University Theses by an authorized administrator of ePublications at Regis University. For more information, please contact epublications@regis.edu.
Use of the materials available in the Regis University Thesis Collection ("Collection") is limited and restricted to those users who agree to comply with the following terms of use. Regis University reserves the right to deny access to the Collection to any person who violates these terms of use or who seeks to or does alter, avoid or supersede the functional conditions, restrictions and limitations of the Collection.

The site may be used only for lawful purposes. The user is solely responsible for knowing and adhering to any and all applicable laws, rules, and regulations relating or pertaining to use of the Collection.

All content in this Collection is owned by and subject to the exclusive control of Regis University and the authors of the materials. It is available only for research purposes and may not be used in violation of copyright laws or for unlawful purposes. The materials may not be downloaded in whole or in part without permission of the copyright holder or as otherwise authorized in the “fair use” standards of the U.S. copyright laws and regulations.
EDUCATIONAL IMPLICATIONS OF HIGH STAKES TESTING

by

Nicole S. Caldwell

A Research Project Presented in Partial Fulfillment of the Requirements for the Degree
Master of Education

REGIS UNIVERSITY

March, 2010

Project Advisor: Gary Upton, Ph.D.
Student achievement has always been a critical concern in education, but it has become a particularly passionate issue since the passage of the No Child Left Behind Act in 2002. As a result of the act, standardized tests have become high stakes, and the futures of students, teachers, and schools have been put on the line. Research indicates that there are many negative implications of high stakes testing. The PowerPoint presentation developed for this project serves as an introduction to these implications and provides ideas for alternative assessment techniques. The target audience for this project is teachers, administrators, law makers, and parents.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>1</td>
</tr>
<tr>
<td>Purpose of the Project</td>
<td>1</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>2</td>
</tr>
<tr>
<td>2. REVIEW OF LITERATURE</td>
<td>3</td>
</tr>
<tr>
<td>History of Standardized Testing</td>
<td>3</td>
</tr>
<tr>
<td>Definition of Testing</td>
<td>3</td>
</tr>
<tr>
<td>History of Standardized and High Stakes Testing Policy</td>
<td>5</td>
</tr>
<tr>
<td>Standardized Testing, Economics and No Child Left Behind</td>
<td>6</td>
</tr>
<tr>
<td>Proponents of NCLB</td>
<td>7</td>
</tr>
<tr>
<td>Opponents of NCLB</td>
<td>8</td>
</tr>
<tr>
<td>Performance Pay</td>
<td>8</td>
</tr>
<tr>
<td>Issues in Standardized and High Stakes Testing</td>
<td>9</td>
</tr>
<tr>
<td>Debate and History Background</td>
<td>9</td>
</tr>
<tr>
<td>Standardized Tests and Socioeconomics</td>
<td>10</td>
</tr>
<tr>
<td>Standardized Testing and the Theory of Multiple Intelligences</td>
<td>11</td>
</tr>
<tr>
<td>Cultural and Language Bias</td>
<td>12</td>
</tr>
<tr>
<td>Gender Bias</td>
<td>13</td>
</tr>
<tr>
<td>Test Anxiety</td>
<td>13</td>
</tr>
<tr>
<td>Inflated Consequences</td>
<td>14</td>
</tr>
<tr>
<td>Test Validity</td>
<td>14</td>
</tr>
<tr>
<td>Testing Environment</td>
<td>16</td>
</tr>
<tr>
<td>Negative Impacts of High Stakes Testing</td>
<td>16</td>
</tr>
<tr>
<td>Testing Ethics</td>
<td>16</td>
</tr>
<tr>
<td>Teaching to the Test</td>
<td>18</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>19</td>
</tr>
<tr>
<td>3. METHOD</td>
<td>20</td>
</tr>
<tr>
<td>Target Audience</td>
<td>20</td>
</tr>
<tr>
<td>Organization of the Project</td>
<td>20</td>
</tr>
<tr>
<td>Peer Assessment Plan</td>
<td>21</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>21</td>
</tr>
</tbody>
</table>
4. RESULTS .......................................................... 22
   Introduction ............................................. 22
   PowerPoint Presentation ............................... 23
   Chapter Summary ....................................... 42

5. DISCUSSION .................................................. 43
   Contribution of this Project ......................... 44
   Limitations ............................................... 44
   Peer Assessment Results .............................. 44
   Recommendations for Further Development ........ 45
   Project Summary ....................................... 45

REFERENCES ................................................... 46
Chapter 1

INTRODUCTION

Student assessment is an integral component of education. However, the use of standardized testing has become a hotly debated topic for a number of reasons. All students learn in their own style, and each teacher teaches in his or her own unique way. Thus, use of the same test for all students does not always fairly represent their level of knowledge and understanding. Conversely, it is impossible to create a unique test for each student. In turn, educators and administration must take this into account and effectively prepare students for standardized tests.

Statement of the Problem

Despite well intentioned purposes, educators, administrators, and law makers have overstated the importance of student performance on standardized tests. According to Nichols and Berliner (2007), test scores do not always provide an accurate representation of how much a student knows. Despite this fact, standardized testing has become high stakes testing. Poor performance can be devastating for everyone from administrators, to teachers, to parents, and most importantly, to students.

Purpose of the Project

The purpose of this project was to show educators, administrators, and law makers the deficiencies of standardized testing and how to avoid them. In addition, this author presents alternative methods for the assessment of student learning. The project
is presented in PowerPoint format and designed to serve as a guide for educators, administrators, and law makers on how to: (a) create meaningful assessments and (b) subsequently, interpret and use test results more appropriately and effectively.

Chapter Summary

In summary, it is this researcher’s opinion that, often, high stakes, standardized testing is detrimental to student and teacher self-efficacy, as the results are so often inaccurate, misinterpreted, or unfairly biased. In Chapter 2, the researcher provides a review of literature that supports the findings that standardized testing is not an effective method of assessment. In Chapter 3, Method, the target audience, organization of project, and plans for peer assessment are presented.
Chapter 2

REVIEW OF LITERATURE

The purpose of this project was to identify the negative consequences of the high stakes testing culture in the United States. Additionally, this author provides important information to educators, administrators, and law makers on how to improve the current procedures in standardized testing. Presented in this literature review are the the history of high stakes testing and the problems that have occurred because of it.

History of Standardized Testing

Over the last 2 decades, standardized testing in the U.S. has become a common practice. Standardized tests turned into high stakes testing with the introduction of the No Child Left Behind Act by the Bush administration in 2001 (as cited in Crawford, 2004). Since then, testing has only become more intense, as have the consequences of the tests.

Definition of Testing

Tests are used as a way to measure something (Gellman 1995). For the purpose of this project, testing measures student learning. According to Gellman, tests should be used as part of a whole in the evaluation of student learning. An analogy used by Gellman explains it best. Standardized test scores are to be viewed in the same way one views weight; a number that provides information, but is not the ultimate judgment. When we step on a scale to weigh ourselves, it is important to take many factors into
consideration, since we can have a relatively low weight and still be unhealthy. Additionally, a healthy weight for a woman is quite different from that of a man, or even from another woman of a different height, body shape, or muscle content. Thus, these measurements should be a tool that is used in combination with other information in order to evaluate our weight (or level of learning as related to test scores).

There are many different types of academic tests (Gellman, 1995). Many standardized tests are considered achievement tests. An achievement test is used to measure a student’s level of knowledge of a particular subject. These types of tests are different from intelligence quotient tests (IQ), which are used to measure intelligence rather than knowledge. Typically, a person’s IQ does not change over the course of time, whereas concepts and information learned can cause changes in a person’s score on achievement tests. There are two categories of achievement tests: mastery tests or norm-referenced tests. Also, mastery tests are referred to as criterion referenced tests and are designed to measure how well an individual has mastered a particular skill. Norm-referenced tests, however, provide norms and indicate how a certain group of students perform on a test. A score on a norm referenced test shows how the test taker performed in comparison to the rest of the group. These types of tests do not measure a student’s mastery of a certain skill, only his or her position in comparison to test taking peers. According to Gellman, either type of achievement test is designed to measure a skill or knowledge, but neither predicts future success or achievement.

Typically, standardized tests are created and published by large testing companies (Gellman, 1995). Often, these tests are given to many students nationwide, which in turn
provides information about the typical performance of students, or the norms. Thus, these tests are used to measure skills and knowledge commonly taught in schools, despite curricular differences. According to Elford (2002), “standardized tests are designed primarily to create summary descriptive statistics based on student responses to questions in an artificial context” (p. xi).

According to Gellman (1995), the results from tests can provide useful information about students. Teachers can determine whether students are confused or have misconceptions about certain concepts. Also, test results can give teachers insight into a student’s strengths and weaknesses. Currently, the staff of many school districts use standardized test results in order to: (a) place students, (b) compare student performance to other districts, and (c) evaluate curricula.

History of Standardized and High Stakes Testing Policy

Nichols and Berliner (2007) traced the recent trend in the U.S. of high stakes testing back to the Cold War. With the Russian progress in the space race came fear that the U.S. was falling behind in many areas, including education. As a result, U.S. officials implemented the Elementary and Secondary Education Act (ESEA; 1965, as cited in Nichols & Berliner), which set higher expectations for student knowledge in reading and mathematics. In order to graduate high school, students in the U.S. were required to pass a minimal competency exam. These tests were very different from the high-stakes examinations that students face currently. They were used to test very basic skills and to ensure that students could read and do basic mathematics at high school graduation. Teachers were not evaluated based on the results from these examinations, and students
were not denied entrance into college because of them as they are today. Eventually, however, many began to criticize the practice because it set expectations for students too low. Thus, the concept of the standards-based achievement test was developed (Nichols & Berliner). Standards-based assessments were created to measure student learning of the particular standards of knowledge set by the department of education in each state. This was the general blueprint for the assessments used in schools today.

*Standardized Testing, Economics, and No Child Left Behind*

Increases in government required standardized testing seem to be correlated with poor performance in the U.S. in areas such as: (a) economics, (b) scientific discovery, and (c) technological advancement. According to Spellings (2005), over 75% of Americans feel that if high schools do not improve, the U.S. will be less able to compete in the global marketplace. This fear has led to educational legislation and reform in recent years. Perhaps the most important piece of legislation in regard to U.S. education was the No Child Left Behind Act (2002, as cited in Crawford, 2004).

Concern over low achievement reports from U.S. schools has steadily increased over the past three decades (Zimmerman & Dibenedetto, 2008). During the Bush administration, the blame was placed on poorly qualified teachers and administrators who led with no incentive to improve. As a result, the No Child Left Behind Act (2002, as cited in Crawford, 2004) was developed to require more accountability in schools and to close the achievement gap by 2014. The Act established: (a) certification requirements for teachers, (b) Adequate Yearly Progress (AYP) reports, and (c) sanctions for schools that did not achieve AYP. These sanctions ranged from increased funding to replacement
of entire school staff. In order to measure AYP, the NCLB required state officials to create standardized assessments of basic skills for students in specific grades in order to receive federal funds for education. As a result, preparation for standardized tests became a large part of classroom instruction. This was an effort to ready students for an end of year examination that, potentially, could bring recognition to a school, its teachers and its administration, or lead to a negative AYP, in the case of poor performance. The Act has caused much controversy in the academic world, which is briefly detailed below (Roller, 2005).

Proponents of NCLB

According to Nichols and Berliner (2007), those who support the legislation of NCLB believe it will put an end to what they term the March of the Lemons. This term refers to the concept that weak teachers are constantly shuffled around schools within a district because administrators struggle to terminate them. According to Nichols and Berliner, often, these lemon teachers end up in the lowest performing schools. With the laws of NCLB (2002, as cited in Nichols & Berliner) in place, the supposed weakest teachers are eliminated when their students underperform on examinations. This same concept applies to underqualified teachers. No Child Left Behind (2002, as cited in Crawford, 2004) requires teachers to have certain credentials, or they are not deemed highly qualified. These requirements are what proponents believe will close the current U.S. achievement gap.
**NCLB Opponents**

Those opposed (Nichols & Berliner, 2007) to No Child Left Behind (2002, as cited in Nichols & Berliner) feel that the test-and-punish aspect of the Act will not be able to overcome the systemic inequalities in race and class that account for the true achievement gap. Nichols and Berliner do not support the act for numerous reasons. They believe it:

(a) treats special education and English language learners unfairly; (b) traumatizes teachers and students due to the stakes attached to scores; (c) ignores out of school conditions such as poverty, lack of medical care, and home situations; and (d) makes decisions about the fate of teachers, students, and schools based on one test.

**Performance Pay**

Another recent trend in education reform is performance pay. In the Denver Public Schools, for example, the movement is called *Procomp* (Denver Public Schools 2010), while North Carolina has implemented what they call a Performance-Based Accountability Program (Nichols & Berliner, 2007). Regardless of the terms, the concept behind these programs is the same: to reward teachers and administrators financially for improved student performance on standardized tests. According to Sacks (1999), critics of such reform claim that this: (a) places too much emphasis on arbitrary examinations, (b) takes away from authentic learning, and (c) creates opportunity for unethical test practices. Sacks noted that proponents of performance pay feel it will eliminate underperforming teachers as well as increase student achievement. It does not appear that the debate will resolve itself any time soon.
Issues in Standardized and High Stakes Testing

Kugler (2003) stated that “while our society is clearly wedded to standardized tests, their use as the sole measure of a school’s worth is fatally flawed” (p. 4). The true purpose of standardized tests is that they be used as one component of student assessment. However, recent increases in the importance of these tests have resulted in many unwanted outcomes. Additionally, the reach of a test scores has expanded to: (a) lifelong labels for students, (b) job loss for teachers, and (c) other negative effects detailed in this section.

Debate History and Background

According to Nichols and Berliner (2007), as early as 1906, officials of the New York State Department of Education felt the concept of high stakes testing was a mistake. The department made the following arguments to state legislature against such practices.

Standardized exams have questionable validity, and general or standard tests are not always accurate representations of a student’s body of knowledge (New York State Department of Education, 1906, as cited in Nichols & Berliner, 2007). It is certainly possible and could be potentially devastating for a well-taught and well-trained student to fail on a high stakes examination. Conversely, it is possible for an unqualified student to score well on an examination due to test deficiencies.

High stakes exams carry unfair and unnecessary weight. When examinations become high stakes, students begin to view them as a final purpose or the ultimate goal of their education (New York State Department of Education, as cited in Nichols & Berliner, 2007).
There are negative effects on instruction (New York State Department of Education, 1906, as cited in Nichols & Berliner, 2007). When a test can make or break a student’s future in life and education, or a teacher’s job security, instruction shifts to a focus on solely test material. This can result in patchy and unsystematic levels of student knowledge which is most certainly not the goal of education. The next sections of this paper describe the difference aspects of the debate in more detail.

*Standardized Tests and Socioeconomics*

Kahlenberg (2001, as cited in Kugler, 2003) argued that the aggregate test scores of a school reflect the socioeconomic status of the school, not the quality of teaching. Scholastic Aptitude Test (SAT) scores are a good example of this, as expensive preparation courses promise an increase of 100 points or more, but only for those who can afford to pay. Duncan, Brooks-Gunn, and Klebanov (1997, as cited in Sacks, 1999) supported the hypothesis that the performance gap on standardized tests has more to do with socioeconomics than with any other factor. Through the use of regression analysis, the researchers isolated many variables to identify which had the most notable correlation with test performance in very young children (e.g., preschool age) on IQ tests. The variables, which had no relationship to score were: (a) gender, (b) birth weight, (c) extent of neonatal care, (d) family structure, and (d) mother’s education level.

Over 80% of the performance gap on IQ tests between Anglo and African American children was attributed to a combination of: (a) family income, (b) the extent to which parents read to their children, (c) whether parents used toys and playtime to develop new skills, and (d) whether children came from a loving home that nurtured
intellectual development (Duncan et al., 1997, as cited in Sacks, 1999). Thus, the research findings suggest that the disparity in test performance lies not in racial differences, but rather in socioeconomics and parental technique. While the specific findings do not explain it, Duncan et al. hypothesized about why the achievement gap is often attributed to racial differences. In this study, African American participants were more likely to report low socioeconomic status than were the Anglo Americans; this may be an explanation of the misinterpretation of test data with regard to racial disparity. African Americans are not less intelligent, but they were more likely to fall into the categories that seem to have a link to low test scores (e.g., low socioeconomic status).

*Standardized Testing and the Theory of Multiple Intelligences*

Gardner (1993) defined intelligence as “the ability to solve problems, or to fashion products, that are valued in one or more cultural or community settings” (p. 7). With this definition in mind, it is hard to imagine a standard test that can effectively measure intelligence. Additionally, Gardner argued that the current state of high stakes testing has a negative effect on students whose inherent intelligence lies outside of U.S. cultural values. For example, students with strong artistic abilities and low scores on reading and mathematics exams are viewed as underperforming. This, according to Gardner, is a result of culture and societal values and is an unfair result of high stakes testing. While multiple intelligence (MI) theory was developed before educational policies like NCLB, the concepts are not consonant with each other. According to Gardner’s MI theory, it is unlikely that a single test can accurately measure all types of learning.
Cultural and Language Bias

A major inefficiency in high stakes testing in the U.S. is the lack of ability to test in languages other than English (Nichols & Berliner, 2007). Depending on the school, district, or even state, there may be anywhere from 0-100% English Language Learners (ELLS) in a school. Despite this fact, these students are required to take standardized tests in English just like their native English speaking peers. This seems unfair for schools with higher populations of ELLs, as ELLs tend to have much lower average scores on high stakes examinations due to the extra factor of difficulty they face when they take tests in a non-native language. As discussed earlier, consistently low scores can have severe consequences for a school. This lack of equity is not accounted for anywhere in high stakes testing and, thus, is an additional argument against its use.

In addition to the unfair advantages that native English speakers have on standardized tests, often, there is cultural bias in these types of examinations as well (Delpit 2006). According to Gardner’s (1993) MI theory, the definition of intelligence is influenced by cultural values. With this in mind, it is hard to justify the use of any one-size-fits-all type assessment in a country often referred to as a melting pot of cultures. Thus, while typically, it is unintended, most standardized tests are biased toward a cultural majority, or at least the cultural majority of the test creators. Delpit’s experience with teaching Alaskan natives revealed a major cultural disparity in the relevance of many examination questions. For example, in a question on an elementary examination, students were asked to choose which form of transportation they would use to get to a hospital: (a) bicycle, (b) ambulance, or (c) helicopter. The Native Alaskan students in
Delpit’s class had never seen an ambulance in their lifetime and, as a result, most chose the third answer, as their response. These students lived in a small fishing village in Alaska, and the nearest hospital was, in fact, a helicopter ride away. All of these students missed a question due to their cultural knowledge and experiences, not because they were not learning. This is an example of students who have actually lived their entire lives in the U.S. Cultural disparity becomes even more notable for students who have recently arrived in the U.S., particularly those from developing countries. Their level of background knowledge is not the same as those who have grown up here. Once again, the result is lower test scores and, in the case of NCLB (2001, as cited in Roller, 2005), school sanctions and possible punishments.

Gender Bias

According to Sacks (1999), in general, males tend to perform better than females on certain standardized tests. However, further research by the College Board (1998, as cited in Sacks) indicated that females, particularly at the university level, consistently earn better grades in school than males. Thus, the use of the Graduate Record Exam (GRE) as a basis for entrance into graduate programs might seem inefficient. Despite this fact, the GRE is still heavily relied upon by colleges in decisions about which students to admit. This is another example of high stakes testing that can result in poorly made decisions about the future of a student’s education based on arbitrary information.

Test Anxiety

Test anxiety is not a new concept. In fact, almost every American has probably experienced some form of test anxiety (Sacks, 1999). In general, the more important a
test, the more anxiety it produces. When a single test can determine whether a student is retained or promoted to the next grade, the stakes become particularly high for students. This is currently the case for many standardized tests. In an effort to increase student efforts on annual tests, many school officials have established a new policy to retain students who do not pass. This kind of pressure can be astounding for students, and the result can be severe test anxiety. Test anxiety can have two particularly damaging results for students: (a) lower test scores and (b) undue stress. When the results from a single test determines whether a student is promoted to another grade, the stakes become unfairly high for students.

**Inflated Consequences**

Test performance can determine the future of a child’s education (Nichols & Berliner, 2007). Often, acceptance into certain programs and schools is based on the results of a test. Scholarships and awards are commonly based on test performance. All of these factors can negatively affect a child’s education. While the use of standardized tests in the case of the No Child Left Behind Act (2001, as cited in Roller, 2005) is aimed at closing the achievement gap, Sacks (1999) argued that, if preschool age children from lower socioeconomic households are refused admittance to high performing schools because of scores on entrance exams, then the gap is just further widened. Further, Sacks argued that a single score on a test should not have so much impact on a student’s future.

**Test Validity**

As stated earlier, the purpose of an academic test is to measure student knowledge. Recent educational legislation (NCLB, 2001, as cited in Nichols & Berliner,
2007) has allowed state officials to create their own standardized tests to measure student achievement in regard to the individual learning standards in each state. The results of tests are partially intended for comparative use, such as the comparison of the reading scores of fourth grade students in Colorado to those in Utah. However, one test cannot comprehensively measure actual differences in students state to state, district to district, or even school to school, for a number of reasons. First, since each state has its own learning standards, the focus of instruction varies by state. Second, when tests become high-stakes, those intent on the attainment of high scores tend to teach what is on the test and in a noncomprehensive way. This is sometimes referred to as rote learning, or even just pure memorization. Thus, high scores from students, who have merely memorized facts that they know will be on a test, does not mean they have a comprehensive understanding of the subject. Additionally, curricular differences can have major effects on student learning. Some curricula may be focused on a certain area of a subject, while another may give a more broad explanation. When students who have a different curriculum take an identical standardized test, differences in scores could be more indicative of differences in curriculum than an indication of student achievement, teacher performance or quality of education (Sacks, 1999). For this reason, critics of high stakes tests, such as Nichols and Berliner, believe they should be used as a general evaluative tool and not considered a comprehensive representation of student learning, unless all other variables (e.g., curriculum, teaching style, learning environment, etc.) are held constant, which is impossible to achieve. Sacks explained that by the very nature of
statistics, standardized test validity is constrained by its ability to comprehensively test
student knowledge of a subject.

Testing Environment

When a student’s standardized test score in North Carolina is compared to a peer
in California, the assumption is that the tests were given under similar circumstances.
However, high-stakes examinations are often given under a variety of test taking
conditions (Sacks, 1999). The human factor in test administration will always provide
for inconsistent testing environments. This may or may not have a statistically significant
impact on results, but it is certainly another way in which the validity of standardized test
scores can be questioned.

Negative Impacts of High Stakes Testing

While Sacks (1999), and Nichols and Berliner (2007) presented reasons why
standardized test scores are not always valid, also, they pointed out that, when these tests
become high stakes, there is a negative impact on education. Students and teachers are
perhaps the most affected, but parents and administrators have felt the effects as well.
However, perhaps the most devastating result of high stakes testing is the impact it has
had on educational ethics. An increase in cheating and teaching to the test was most
certainly not the intended result, but it has become increasingly common (Nichols &
Berliner).

Testing Ethics

Nichols and Berliner (2007) asserted that the scores from high-stakes testing
cannot be trusted due to distortion and corruption. When test results carry great
importance, the likelihood of cheating increases. Hundreds of schools and educators have been caught either blatantly skewing results or doing so in a more subtle fashion. While many have been caught, most likely, even more have not. In a national survey of teachers conducted by Edmonson (2003, as cited in Nichols & Berliner), it was found that approximately 10% admitted that they provided hints about test answers to students and pointed out mistakes. Another 15% of teachers reported that they had provided extra time to students, and 5% admitted to the provision of instruction during test time. Such information provides little hope that any given test result is accurate, since there is no way to know which results are altered. Nichols and Berliner cited numerous documented examples of cheating on standardized tests, such as:

1. school counselors in Tennessee locked themselves in offices to erase stray marks,
2. teachers in North Carolina shared copies of examinations with students before test day,
3. Wisconsin teachers had students memorize the sequence of answers on multiple choice tests, and
4. a principal in Maryland coached students and blatantly gave answers to test questions.

These are just a few examples of unethical testing practices that have taken place in the past. With cheating such a strong possibility, Nichols and Berliner questioned the validity of any test scores.
Aside from obvious examples of cheating, high stakes testing can result in many other unethical actions by educators and their administrators (Sacks, 1999). For example, before states were required to test a minimum of 95% of the school population, school staff were allowed to exclude the scores of students with identified learning disabilities. As a result, special education referrals increased in an effort to eliminate lower scores.

Nichols and Berliner (2007) referred to an incident in Florida that they noted is not uncommon. When the NCLB (2001, as cited in Nichols & Berliner, 2007) linked federal funds to school performance on annual tests, often, low scoring students were suspended for the smallest infraction during test week, while high scoring students involved in similar or worse behaviors were given lighter punishments.

*Teaching to the Test*

With jobs at stake and school reputations on the line, the temptation to teach to the test can be high. Sacks (1999) noted that the extent to which students receive instruction focused on the attainment of higher test scores cannot be measured or accounted for in the consideration of test results. Some school staff spend excessive funds on expensive test curricula in an effort to boost scores, while others may take a more authentic approach to teaching and hope it proves successful. Regardless of which method is right or wrong, it changes and skews results, again, the validity of test scores are questioned. The unfortunate repercussions of teaching to the test can be fragmented knowledge of facts and little understanding of concepts, which is not beneficial for students. Also, teaching to the test can decrease student engagement in learning, a side effect that is most certainly not advantageous to education.
Chapter Summary

The concept of evaluating human learning and performance is not new. However, the recent trend to measure knowledge, predict academic success, and assess teacher performance based on a one-size-fits-all examination has raised numerous questions about the validity of standardized tests. It is necessary and important to monitor student learning and understanding but, at the same time, it is a very difficult and complex task. Many different theories about how tests should be developed and used have created a major controversy in educational reform. The objective of this project is to further educate teachers, administrators, and law makers on these issues and provide alternative ways to assess student performance. In Chapter 3, the author describes the method used to create the PowerPoint presentation.
Chapter 3

METHOD

The purpose of this project was to create a PowerPoint presentation to educate readers and provide alternatives for the current standardized testing culture. The author discusses current issues with high stakes testing and provides suggestions for change. Additionally, the author provides information on how to: (a) create meaningful student assessments, and (b) interpret and use test scores in appropriate and effective ways.

Target Audience

The target audience for this project is teachers, school administrators, educational lobbyists, and law makers. While not created purposely for parents, the project also provides helpful information so that they too may better understand the use and results of tests.

Organization of Project

The project takes the form of a PowerPoint presentation. First, the issues with high stakes testing are identified. Subsequently, the author offers suggestions for change in standardized testing practices. Finally, the author provides information and support for teachers on how to effectively use the test results they have available while the current high stakes testing requirements are still in place.
Peer Assessment Plan

The project was informally reviewed by three teachers and one administrator. The PowerPoint file was sent to each peer for review and informal feedback on project content and clarity. Responses were given verbally to the project author. Their feedback is discussed in Chapter 5.

Chapter Summary

High stakes testing has become a hotly debated topic in education. Despite the controversy, the practice is still widely used. This author applies knowledge gained from research of the pertinent literature and presents, in PowerPoint form, a critical evaluation of high stakes testing and makes suggestions for improvement and change. In Chapter 4, the author presents the PowerPoint slides.
The testing culture of the United States has become more intense since the passage of No Child Left Behind in 2002. While on their own, standardized tests serve a purpose, there are many issues with the implementation of high stakes tests. After careful research, this author created a presentation meant for teachers, administrators, parents and law makers citing the many downfalls of standardized tests and high stakes. In addition, thoughts for how to improve the use of standardized tests in education were outlined along with alternate assessment techniques. The project is in the form of a Power Point document and is presented in this chapter.
Educational Implications of High Stakes Testing

Nicole Caldwell
Master of Education Student
Regis University
“Standardized tests are designed primarily to create summary descriptive statistics based on student responses to questions in an artificial context.”
- Elford, 2002 p. xi
Intended purposes of standardized tests

- To be used as part of a whole to evaluate student learning and understanding
- To indicate areas in which students are lacking knowledge
- To aid teachers in creating lesson plans
- For initial placement, such as Advanced Mathematics or 2nd year Spanish

- Scores as part of a whole
- Student should never be permanently placed and labeled due to one score on a standardized test.
- Other factors, such as classroom performance, are important to consider when placing students in classes and levels.
- Students who get extremely nervous while speaking Spanish may not perform well on a fluency test, but should not be placed in a beginning class if they are strong in reading and writing the language.
Standardized tests became high stakes when:

- The SAT/ACT/GRE etc., became a highly influential element in college applications
- NCLB was written into law
- Teachers began to lose jobs
- Students were permanently labeled (e.g., Special Ed)

- When a single standardized test has massive repercussions, it becomes a “high stakes” exam.
- When No Child Left Behind was written into law, high test scores became a way for schools to receive funds
- When teachers lose jobs because of student tests scores alone, the stakes become high.
- When children are labeled as “special ed” because of a test, they are generally permanently marked as such, which can have many unfortunate outcomes (e.g., lowered expectations, social ramifications, assumed inability to learn, self esteem issues, etc.).
Standardized tests were NOT meant to be indicative of:

- Future success
- Ability to learn
- Quality of teaching
Issues with standardized test validity

- Socioeconomics
- Cultural Bias
- Language Bias
- Gender Bias
- Testing Environment
In a 1999 study, Duncan, Brooks-Gunn, and Klebanov found that standardized test scores were more related to socioeconomic status than any other factor.

Implications: Students of lower socioeconomic status are deemed underperforming, quality of these students’ teachers is questioned, students are viewed as a liability by schools.

- An extremely interesting and revealing study conducted in 1999 found that, more than any other factor (e.g., race, mother’s level of education, family structure, birth weight, extent of neonatal care, etc.), family income was the most accurate predictor of performance on standardized tests.
- If NCLB continues as law, students of low socioeconomic status will continue to underperform and thus lose more money for their schools. Schools with students of high socioeconomic status will continue to get more money.
Most any exam will contain cultural bias of some sort. It is nearly impossible to eradicate all cultural bias due to the fact that culture is so varied and subjective. Students come from a variety of backgrounds and thus have a variety of background knowledge. Test makers have to assume certain basic knowledge when creating tests and they tend to use their own experiences as a basis, which is not objective or fair. Implications: Students of cultural minority score lower because they have different background knowledge and as a result, the quality of their school and teaching is questioned.

- When this bias begins to affect the outcome of the test scores, the test becomes invalid, UNLESS the test is measuring culture, which is not usually the case.
English Language Learners have a clear disadvantage when taking a standardized test in English.
Even different parts of the U.S. have differing dialects.
Implications: Schools with more ELLs have lower scores and thus reputations as “underperforming,” teachers and administrators lose jobs, schools are closed, ELLs become viewed as a liability.

- Think about a time when you were learning a new language. Was it hard? Was it ever nerve-wracking? Have you ever moved to another country and had to take a test in that other language? A test that was created for people native to that country and language? If you have, you probably know that it is a very difficult thing to do. If you have not, you can probably imagine what it would be like.
- While we can provide accommodations for these students, such as more time or reading them the test, we cannot tell them what a word means or explain an idiom.
Gender bias

- In general, males tend to perform better than females on certain standardized tests; however, in the college setting, females consistently maintain higher grades than males (College Board 1998, as cited in Sacks, 1999).

- Implications: more males accepted to college programs because of high test scores.
Testing environments will always differ. There is no way to account for the human aspect involved in testing, which can have an effect on scores.

Implications: A fire alarm goes off in the middle of the exam. Student concentration may or may not be affected, but there is no way to know for sure. Scores decrease and teachers are blamed?

- Other examples of test environment distractions: room temperature, anxiety of other students, etc.
Negative implications of high stakes testing

- Test Ethics
- Teaching to the Test
Test ethics

- In a national survey conducted by Edmonson (2003, as cited in Nichols & Berliner, 2007) of teachers, it was found that approximately 10% admitted that they provided hints about test answers to students and pointed out mistakes. Another 15% of teachers reported that they had provided extra time to students, and 5% admitted to actually teaching during test time.

- Implications: Test scores are invalidated, students are exposed to immoral behavior by teachers.
When the results of student scores can make or break a teacher’s career, the tendency to spend more time on test material than other subjects is strong.

Implications: Instructional time is turned into drills, memorization activities and test preparation. Students lose interest in learning and have fragmented, fact based knowledge instead of conceptual understanding of subjects.
What can be done now

- Use standardized tests in the way in which they were meant to be used: as a part of a whole.
- Understand what test scores say, maintain skepticism of sweeping statements about entire groups of students and teaching staff based on test scores.

- Standardized testing will most likely not be eliminated completely any time soon.
- However, there are ways in which these scores can be used that are efficient and meaningful.
- For example, when a standardized test score is part of a student portfolio, and not the sole means of assessing a student’s level of learning, it can be useful.
Work to take bias out of test questions. This may mean that standardized tests differ from state to state, district to district, and even school to school.

- Develop new systems of evaluation for both students and teachers. Portfolios, packets, work samples, etc.
- Shift thinking. Education is not cut and dry, but standardized tests are. It is impossible to measure and represent something as complex as learning using only numbers and statistics. It is just not that simple. “Educating young people is not like making motorcars” (Robinson 2010, as quoted in Sutter 2010, p. 1).

The most important change should be in the way we regard these tests.

One-size-fits-all examinations should never be the end-all, be-all of a student or teacher’s future.

Education is a far too complicated task to be reduced to simple numbers.

A number or score can never represent the entirety of a student’s knowledge and understanding.
Alternatives

- Authentic Assessment
- Teacher judgment
- Portfolios
- Authentic Assessment is used in classrooms on a daily basis.
- This would require no big changes in teacher work and would eliminate the rote learning that happens when we teach to a test.
- Authentic Assessment also eliminates many test skewing factors such as test anxiety and bias.
Teacher Judgment

- High school grade point average is the number one predictor of success in college (Elford, 2002).
- Teachers tend to know the abilities of their students very intimately. Report cards are often standards based and thus provide a detailed description of what students do and do not know.

- Already used on a daily basis
- This is not to say that teachers do not utilize standardized test scores to come to these conclusions, but teachers tend to take other indications of learning into consideration as well, which can be a more comprehensive way of assessing.
A recent study done in the United Kingdom (Marshall, 2008)) showed teachers and administration a way of assessing students through the use of portfolios. In the study, there were standards and goals set for students to achieve. Once they met these certain steps, they were moved on to the next “level” (similar to grades). Each standard had a set rubric and students would compile work in a portfolio to be reviewed by a teacher. The teachers did not know the identity of the student whose portfolio they were assessing and thus had no level of bias. While the practice is still in its beginning stages, it appears to be a useful and progressive way of assessing student learning.
“While our society is clearly wedded to standardized tests, their use as the sole measure of a school’s worth is fatally flawed” (Kugler, 2003, p. 4).
Chapter Summary

The PowerPoint presentation in this chapter provides teachers, administrators, law makers, and parents with research regarding the implications of high stakes testing. This presentation also offers suggestions for ways to improving and changing the negative impacts of such testing with alternative assessments. In Chapter 5, the author provides a final discussion of the completed project.
Chapter 5

DISCUSSION

In recent years, standardized test scores have become a major emphasis in education and education reform. The results of having a high stakes testing culture in schools have not always been positive and in fact have had some powerful and devastating consequences for students, teachers, and schools. Alternatives to using standardized test scores as the sole means of evaluating students and teachers have started to develop, but they are still, for the most part, disregarded. Without some change in the testing culture, the possibility of producing students with amazing test abilities and few problem solving skills will only increase. It is an issue that needs to be faced.

Contribution of this Project

The purpose of the project was to create a PowerPoint presentation to inform and educate teachers, administrators, law makers, and parents about the implications of high stakes testing and to suggest ways to change the current system of evaluation. The contribution this project makes to the educational community is important in that it presents facts about a topic which in turn allows individuals to form their own opinions. In addition, the project not only highlighted the controversy surrounding high stakes tests, but it also suggested ways to improve the situation with specific techniques.
Limitations

A limitation of this project is that many of the alternatives to standardized tests have not been studied long term. It is not known how effective these other types of assessments might be or if they will contribute to improved student performance.

Peer Assessment Results

The presentation was reviewed by a group of three teachers and one administrator, all of whom are also parents of students affected by high stakes testing. The assessors were asked to evaluate the presentation in terms of content and clarity. Their feedback was received informally through verbal communication. Assessor comments were as follows:

1. I would possibly add something about not just teachers being fired, but entire schools being shut down. As you know high stakes testing can have a huge impact on an entire community, not just a teacher.

2. I totally agree with your premise that standardized testing is not evil in itself; it is very useful to compare students within your state or nationally. It can be a great piece of data in an entire body of evidence. You are right that what is not good about standardized testing is making them "high stakes." They cannot be used as the ONLY indicator of an effective teacher, good school, or of a child's abilities. They are also definitely not useful in determining how to modify teaching, which is what assessments are supposed to be. For one thing you don't even see the results until the next school year and to be honest, I
don't think the next year teacher uses them unless it is for test preparation for the following year.

3. My biggest problem with testing is that the students have no personal stake in the outcome. Their lives are not affected at all. Especially by high school (when the scores really drop) the students don't care how they make their school look.

The next set of comments are regarding the clarity of the presentation:

1. I would add the year or years when testing became so important on slide 4.

2. I would include where you got your information on what the tests were not intended to do (slide 5).

Changes were made to the project as a result of some suggestions made by the evaluators. In particular, the inclusion of school closure as a repercussion and negative impact of low test scores was added in the implications section on slide 9. The author chose not to add the specific dates and reference information as suggested in slides 4 and 5.

Recommendations for Further Development

Further study of alternatives to standardized test scores as assessments is necessary to really determine the effectiveness of such techniques. The extent to which students are motivated by test scores or alternative assessment is another area of recommendation for further research. Also, an entirely new system of teacher and school evaluations would need to be developed if the use of standardized test scores was to be eliminated.
Project Summary

This project presented the target audience (teachers, administrators, law makers, and parents) with an introduction to the educational implications of high stakes testing. The project also presented alternatives to the use of standardized high stakes tests, such as portfolios and authentic assessment techniques.
REFERENCES


Kugler, E. (2003). Debunking the middle-class myth: Why diverse schools are good for all kids. Lanham, MD: Scarecrow Education.


