High-stakes Standardized Testing in Schools

Samantha Destremps
University of Rhode Island, sdestremps@gmail.com

Follow this and additional works at: http://digitalcommons.uri.edu/srhonorsprog

Part of the Curriculum and Instruction Commons, Disability and Equity in Education Commons, and the Educational Assessment, Evaluation, and Research Commons

Recommended Citation
http://digitalcommons.uri.edu/srhonorsprog/473
High-stakes standardized testing in schools

Samantha Destremps

University of Rhode Island
Standardized tests have been in use for decades, but high-stakes testing policies were mandated under No Child Left Behind (NCLB) in 2001 (Boehner, 2002). High stakes testing is defined as having “significant consequences for students, teachers, schools, and school districts” (Haladyna, 2006). As a result of NCLB there has been a movement in which school districts, schools, and teachers are being held accountable for student achievement. As a way to measure student achievement, states have implemented test-based accountability using standardized tests and put in place high-stakes consequences in an attempt to increase motivation of students, teachers, and schools (Haladyna, 2006).

Prior to NCLB, the Educate America Act (1994), also known as Goals 2000, implemented a federal grant program in which states that wanted to receive funding had to develop standards and include a way to measure whether or not students are meeting the standards (O’Neill, 2001a). Goals 2000 required states to include all students, including those with disabilities, and implement necessary accommodations and adaptations for the testing of all students (O’Neill, 2001a). Title I of Improving America’s Schools Act requires states to submit annual progress reports every year. The intentions of the Educate America Act and NCLB were to improve schools across the country; states wanted to make sure that the diploma was meaningful by making sure that students graduated knowing certain information and having specific skills. (Marchant & Paulson, 2005).

Along with laws implementing standardized testing, there has also been legislation requiring all students, including those with disabilities, to participate in testing (O’Neill, 2001a). Section 504 of the Americans with Disabilities Act prohibits discrimination based on disabilities and, as such, individuals are qualified to participate in testing.; it also requires schools to have accommodations for students during testing (O’Neill, 2001). The primary federal law regarding
students with learning disabilities and education is the Individuals with Disabilities Education Act 1995 and amendments; this law requires each student with a disability to have an Individualized Education Plan (IEP) (O’Neill, 2001). In the IEP there must be goals and measures written down for each student and it must include to what extent each student will participate in testing and what accommodations they will have (O’Neill, 2001). This law also requires states to have data on how students with disabilities perform on tests but leaves it up to the states to determine if and what consequences there are for performance (O’Neill, 2001).

Over time, test scores began to be used to make decisions regarding students, teachers, schools, and districts. The results of tests are used to inform policy, help to determine where resources are needed and to compare groups of students to see where there are gaps (Haladyna, 2006). Public opinion of schools are also informed by the released test results. In some states there are severe, negative consequences for both students and teachers if students do not perform well on tests (O’Neill, 2001). Test scores can also be used to make decisions such as student grade promotion and graduation, teachers’ pay and continued employment, and intervention and funding of schools and districts (Haladyna, 2006).

Since the 1970’s, there have been some consequences tied to performance on exit exams due to the emphasis on reform. Based on standards, there has been an increase in the number of states requiring students to obtain a passing score on standardized tests in order to graduate (O’Neill, 2001a). The incentives and consequences tied to standardized test scores are meant to make change happen very quickly (Herman & Golan, 1990). In order to receive funding states must implement tests and show students are making progress (Robinson & Aronica, 2016).

One of the important tasks when implementing standardized testing is to make sure the test is valid; Haladyna (2006) defines validity as the whether or not the interpretation of a test-
score is an adequate representation of what the test is supposed to measure (2006). Many researchers have found that there can be many problems with the validity of standardized tests that can compromise the results (Haladyna, 2006; Herman & Golan, 1990; Marchant & Paulson, 2005; Robinson & Aronica, 2016). Haladyna (2006) found that there are two types of problems with the validity of standardized tests, content irrelevance and content underrepresentation. Content irrelevance includes factors that increase or decrease scores for some students inaccurately or systematically (Haladyna, 2006, p. 34). Content under representation includes “flaws in the design of the test that fail to evaluate its full range of content and cognitive behaviors” (Haladyna, 2006, p. 34). Herman and Golan (1990) found that teachers are not confident in the validity of standardized tests; there are many factors that contribute to a student’s performance on a standardized test that a teacher cannot account for, including the appropriateness of the test and the self-confidence and motivation of students.

A major concern of researchers, educators, and policymakers is that the test adequately measures what is being taught and that the score a student receives accurately shows how much they know. In order to do so, the tests and scoring of tests must remain consistent and be documented over time in order to show that the test is a valid measure of student achievement and progress (Haladyna, 2006). One concern is the idea of teachers teaching to the test and leaving out information that is not tested; when under a lot of pressure to increase scores this practice becomes necessary, to some extent, although it may bias test results (Haladyna, 2006; Marchant & Paulson, 2005). The content of the tests should match the curriculum used by teachers and be an accurate reflection of what students learn. Researchers have found that when under pressure from principals, superintendents and states, teachers change their curriculum and
teaching methods in hopes of increasing student test scores (Au, 2007; Herman & Golan, 1990; Robinson & Aronica, 2016).

An article written by Bianca Tanis (2015), a special education teacher and mother in New York, discussed the ways in which standardized testing can take away autonomy from teachers. Teachers have less opportunity to use their own methods of measuring student achievement, because they are trying to get in as much information that might be on the standardized test, at the expense of using their own assessments and making sure students know the information before moving on (Tanis, 2015). Other studies have found similar results in regards to the effects on teachers from standardized tests, which directly influences the students (Herman & Golan, 1990; Robinson & Aronica, 2016). Herman and Golan (1990) found that the effects of standardized testing on teachers are influenced by whether or not the teacher believes they are responsible for student’s scores or whether they believe it is out of their control. This study also found that teachers feel more pressure if they feel they are being evaluated based on their students’ scores (Herman & Golan, 1990). There was also an increase in anxiety and stress of teachers if the scores were made public and they felt more pressure from their principals (Herman & Golan, 1990).

A study by Au (2007) found that tests control three areas of curriculum and teaching; content, formal, and pedagogic. The ways in which the curriculum is altered depends on a number of factors including the test format, design of the test, subject matter, and teacher characteristics (Au, 2007). Existing curricula can be expanded or restricted based on the standardized tests (Au, 2007). This can include narrowing the focus onto certain topics, cutting out topics altogether, or expanding topics into other areas. (Au, 2007). The formal area of curriculum is how the knowledge is broken down and taught in relationship to other parts of the
High stakes standardized testing in schools

curriculum (Au, 2007). The study found that in many cases teachers broke down material into small chunks based on how they are tested rather than integrating pieces of material together (Au, 2007). The pedagogic area of curriculum relates to the teaching strategies used to transmit the knowledge: teacher-centered or student-centered teaching strategies (Au, 2007). Au found that many of the participants changed their instruction strategy to be more teacher-centered in order to transmit the large amount of information needed for the test in the fastest way (Au, 2007). A small number of participants in this study found that their curriculum went in the opposite direction by expanding the content, integrating topics more, and using more student-centered instruction as a result of standardized testing (Au, 2007).

A study by Herman and Golan (1990) found that standardized testing influences teaching and curriculum. When looking at whether or not tests affect teachers decision making, they found mixed results; some studies found that tests have little effect on their decision making, while other studies found that tests helped teachers make decisions (Herman & Golan, 1990). Herman and Golan (1990) found that when high-stakes were tied to standardized test scores, the curriculum was narrowed; teachers taught to the test and left out assignments and teaching strategies that were not covered on the exam. Teachers have been found to use more worksheets and practice tests for assignments that end up de-emphasizing reasoning and critical thinking (Herman & Golan, 1990; Marchant & Paulson, 2005). Standardized tests typically ask students to respond in only a couple of formats, usually multiple choice and essay responses (Robinson & Aronica, 2016). Herman and Golan (1990) found that teachers use assignments that are in the same format as the test, particularly in regards to writing assignments. This study also found that teachers spent a large amount of time teaching test-taking strategies and that the time spent increased as they got closer to the test date (Herman & Golan, 1990; Robinson & Aronica,
Robinson and Aronica (2016) found that standardized testing can cause teachers to leave out learning experiences that would promote creativity, problem solving, and collaboration, and critical thinking skills in order to fit in as much content from the test as possible.

Another concern regarding standardized testing is the scoring and administration of the tests. Haladyna (2006) found that changes in the administration of tests alter the validity of the tests. This study found that some schools “clean up” student answer sheets to make sure changed answers are erased properly and that there are no stray pencil marks; this becomes a problem when some schools do this and others do not (Haladyna, 2006, p. 39). This practice, and other similar practices, systematically influence test scores for some students and not others.

Test scoring can also impact a student’s performance, beyond the control of the student or the teacher. There have been cases in which large numbers of tests were scored incorrectly; one case is from the Scholastic Assessment Test (SAT) in which too much moisture in the testing sheets caused inaccurate results (Haladyna, 2006). There are also problems regarding the inter-rater reliability, particularly when grading open-response questions. Open response questions typically allow for more valuable information regarding student knowledge and understanding to be gathered, but is also more expensive and less reliable to grade (Robinson & Aronica, 2016).

Due to the increased pressure put on students, teachers, and schools to perform well on tests, there is a concern over cheating, which would impact the validity of test scores (Haladyna, 2006). As a result there is an increase in the amount of time and energy spent to increase the security of the test and the detection of cheating (Haladyna, 2006).

Changes in curriculum and teaching strategies have a direct influence on student learning. In addition to influencing what knowledge and skills students accumulate over the course of their education, there can also be emotional and social consequences as a result of the testing process.
and their performance on the test (Robinson & Aronica, 2016; Tanis, 2015). Tanis (2015) reports having to call the school psychologist to counsel students who have either shut down or started to cry during the testing process. Students and teachers report that the atmosphere of the school changes during testing and the time leading up to it; in some cases the testing experience can be damaging (Robinson & Aronica, 2016; Tanis, 2015). Researchers also report changes in students self-esteem and self-confidence as a result of testing; when stakes such as grade promotion and graduation are influenced by test scores students face increased pressure to do well or face serious consequences (Robinson & Aronica, 2016; Tanis, 2015).

Another effect on student learning is the emphasis on the subjects tested (mostly English and math) and de-emphasizing subjects that are not tested (art, gym, foreign language, music, and sometimes science and social studies) (Robinson & Aronica, 2016). Students that require extra help to improve scores are typically pulled out of these classes and end up losing time in these subjects as a result.

Certain student populations are more vulnerable to the negative effects of standardized testing and consequences including students with learning disabilities, English-language learners, and students from low-income or impoverished homes (Herman & Golan, 1990). Students with learning disabilities face many of the same challenges with standardized testing as general education students, but the effects can be even greater; an example would be not properly scaffolding information in an effort to get as much information in as possible before the test (Tanis, 2015) Students with learning disabilities have extremely diverse learning needs and are required to take the same “one-size-fits-all” test that all students take (Tanis, 2015). Another problem with the testing is with what they measure; standardized tests measure what students do not know, rather than how much they have progressed over time. This is particularly problematic
High stakes standardized testing in schools

for students with learning disabilities because of their diverse learning needs and goals; students with documented learning disabilities have an IEP, which states their goals and measures, which may or may not line up with the goals of the state and performance on standardized tests (O’Neill, 2001; Robinson & Aronica, 2016).

Students with learning disabilities have accommodations available to them and put in their IEP; accommodations are supposed to level the playing field for students with learning disabilities (Bouck, 2013; O’Neill, 2001). Standard accommodations include extra time, small group, read a-louds, preferential seating, and frequent breaks (O’Neill, 2001). Some states have an alternative assessment available to some students with disabilities; all states allow for students to retake a version of the assessment multiple times (O’Neill, 2001). In many cases the accommodations available do not adequately level the playing field and accurately assess what students know (O’Neill, 2001). There are also students with mild intellectual disabilities that do not receive accommodations and take the general standardized assessment (Bouck, 2013). Including all students in standardized testing was meant to hold schools accountable for holding students to high standards; however, in many cases standardized tests do not measure learning and progress towards goals (Robinson & Aronica, 2016).

A study by DeBono and colleagues (2011) looked at how students with attention-deficit/hyperactivity disorder (ADHD) perform on different measures of written expression. This study found that the relationship between learning disabilities and scores on standardized tests is very complicated (DeBono et al., 2011). There are high rates of children with ADHD also having some type of learning disabilities; many students with ADHD have difficulty academically (DeBono et al., 2011). Written expression involves many cognitive processes that affect students with learning disabilities such as “working memory, processing speed, language, fine motor
ability, and reading efficiency” (DeBono et al., 2011, p. 1403). In many cases, accommodations available such as extra time do not adequately level the playing field for students with disabilities.

The standards of each state’s curriculum and the consequences for not obtaining a passing score are left up to the individual states; some states have more rigorous standards and severe consequences than others (O’Neill, 2001). Not all states require students to pass a standardized test in order to graduate high school; some states that do require a passing score offer alternative paths to graduation or another type of diploma (O’Neill, 2001). There have been court cases that have challenged whether or not states can require a passing score for students to receive a diploma and have set regulations for doing so (O’Neill, 2001). The case Debra P. v Turlington looked at the Florida’s law requiring students to pass a “statewide minimum competency test” in order to obtain their diploma (O’Neill, 2001, p. 197). The court ruled that the student’s property interest was at stake and that the state could not impose that restriction without giving the students enough notice and time to prepare (O’Neill, 2001). Students need time to prepare, states and districts need time to prepare and administer tests, and there needs to be time to correct any problems and set a cut-off score; expert evidence proposed this would take at least four to six years (O’Neill, 2001).

It is important to look at what is at stake when imposing severe consequences for not passing a standardized test; there are academic, social, and emotional effects for students that are denied a high school diploma (O’Neill, 2001). Students who leave school without a diploma are at a disadvantage compared to their peers with a diploma (O’Neill, 2001). Marchant and Paulson (2005) found that states with requiring a passing test score had lower graduation rates.
State and federal policies are constantly changing in an attempt improve student learning. The Every Student Succeeds Act (ESSA) enacted in 2015 is the most recent federal legislation regarding education (Andrejko, 2015). ESSA maintains annual testing, but allows states to determine their own standards and what the consequences of the test scores will be (Andrejko, 2015). This act leaves more room for other measures of assessing students, moving away from using one score to make important decisions (Andrejko, 2015). Evaluation of teachers is left up to the states and schools will not be punished for not meeting progress goals (Andrejko, 2015). As ESSA is put in place the practical effects of this legislation will be seen.

Throughout the literature there are many suggestions put forth to improve upon the current system of accountability (Curwin, 2014; Robinson & Aronica, 2016). The current system measures what students have not learned rather than measuring the progress they have made. Standardized testing is a short-term fix to address problems in education; instead there need to be solutions that address the root of the problems (Robinson & Aronica, 2016). Robinson and Aronica (2016) suggest ways to make the current system of testing more fair, such as allowing for more individualized accommodations to students. The researchers also propose better ways to assess student learning and achievement such as using research-based assessments and looking at long-term achievement such as success in college and/or employment (Robinson & Aronica, 2016).

Robinson and Aronica (2016) suggest a large-scale assessment, similar to a portfolio system, in which students and teachers set goals, collect pieces of student work, and analyze progress made over time (Robinson & Aronica, 2016). This system involves students in the process of assessment and tracking progress and recognizes the progress students make over time, rather than documenting where there are deficits (Robinson & Aronica, 2016). This method
of assessment also leaves more room to catch problems early; the testing system usually finds problems when it is too late (Robinson & Aronica, 2016). Curwin (2014) proposed a way to increase student motivation by acknowledging effort and including that in their grade. Curwin (2014) asserted that in order to do this effectively teachers need to accurately measure effort to be fair to all students. Examples of ways to measure effort include counting grades at the end of a term more than those at the beginning in order to account for improvement, counting if a student seeks help or helps other students, and allow the opportunity for extra work and redoing assignments (Curwin, 2014). Each of these methods of learning assessment deserve further study as potential means to expand assessment options, increase their meaningfulness, and alleviate the disproportionate emphasis and consequences associated with current standardized testing.

In conclusion, high stakes standardized testing can in some cases, improve schools. However, in many cases high-stakes standardized testing can cause more harm than good for both teachers and students. There is the potential for many problems with the validity of standardized tests, which can cause scores to be inaccurate representations of student learning. If high-stakes are attached to test scores there is the potential to cause lasting harm to students academically, socially, and emotionally. Although there is research showing mixed results on the effects standardized testing on learning experiences, there is other research to show that the pressure teachers feel can, in many cases, cause drastic changes in teaching and curriculum that does not improve student learning (Au, 2007). I believe students’ education should account for the extremely diverse learning needs of students rather than having all students take a one-size-fits-all assessment. Thus, in order to do so, the assessment system should look at progress and gains rather than deficits in student learning. Comparing students to themselves rather than other
High stakes standardized testing in schools

students can promote a more productive learning environment and increase individualization of education.
Works Cited


