



INFORMATION CAPSULE

Research Services

Vol. 1008
January 2011

Christie Blazer, Supervisor

UNINTENDED CONSEQUENCES OF HIGH-STAKES TESTING

At a Glance

High-stakes testing is one of the most controversial issues in American education. Advocates contend that these tests encourage students to work harder, provide teachers with a stronger understanding of students' strengths and weaknesses, and allow educators to target failing schools for extra help. Critics claim that they narrow and distort the curriculum, hold students and teachers with inequitable resources to the same standards, and solidify class and ethnic disparities. This Information Capsule reviews research conducted on the unintended consequences of high-stakes testing programs, such as narrowing of the curriculum, higher levels of student test anxiety, and increased pressure on teachers. In addition, high-stakes tests have been found to have a disproportionately negative impact on low-performing, low-income, and minority students. Although the majority of unintended consequences are negative, researchers have found that high-stakes tests have some positive effects on education, including increased teacher professional development, better alignment of instruction with state content standards, more effective remediation programs for low-achieving students, and increased use of data to inform instruction. The research is mixed on the impact of high-stakes testing on dropout rates, students' levels of academic achievement and motivation, and on the consequences of publishing test scores. This report also includes a brief review of studies that have examined the full costs of high-stakes testing.

High-stakes testing is one of the most controversial issues in American education. While some policymakers view testing as an important component of the school reform and improvement process, others see it as a threat to the quality of teaching and learning (Mitchell, 2006; Burger & Krueger, 2003; Westchester Institute for Human Services Research, 2003).

Tests are called high-stakes when they are used to make decisions about students, teachers, schools, and/or districts. Examples of tests associated with high-stakes for students include those used to make decisions about high school graduation or grade promotion. Tests that are high-stakes for educators include those used to make decisions regarding teachers' jobs and pay. High-stakes tests for schools and districts often determine school funding levels and guide school restructuring efforts (von der Embse, 2008; FairTest, 2007; Moses & Nanna, 2007; Marchant & Paulson, 2005; Pedulla et al., 2003; Westchester Institute for Human Services Research, 2003; Mulvenon et al., 2001; Lewis, 2000).

Research Services

Office of Assessment, Research, and Data Analysis
1500 Biscayne Boulevard, Suite 225, Miami, Florida 33132
(305) 995-7503 Fax (305) 995-7521

Prior to the 1970s, large-scale tests were used to provide teachers with information about their students, but rewards or sanctions were rarely associated with performance. Policymakers started using test results to make decisions about individual students during the minimum competency testing movement of the 1970s. The emergence of formal accountability systems in the 1990s resulted in testing programs that provided incentives or sanctions to both students and schools on the basis of test scores. With the passage of the No Child Left Behind Act of 2001, states began attaching increasingly higher stakes to tests. Today, American students are tested with greater frequency than at any other time in the history of the United States (Elbousty, 2009; Jacob, 2005; Nichols, 2003; Stecher, 2002; Steeves et al., 2002; Wright, 2002).

Advocates of high-stakes testing contend that these tests provide the pressure that is needed to encourage students to work harder and teachers to adopt more effective practices. They maintain that high-stakes tests provide students with information about their own knowledge and skills and teachers with a stronger understanding of individual students' strengths and weaknesses. Proponents also point out that when schools are held accountable for student achievement, high-performing schools can be rewarded and failing schools can be targeted for additional assistance and resources (Elbousty, 2009; Perkins & Wellman, 2008; Jones & Egley, 2007; Barnes, 2005; Marchant & Paulson, 2005; Stecher, 2002; Wright, 2002; American Psychological Association, 2001).

Critics of high-stakes tests, on the other hand, claim that they increase students' risk of educational failure, hold students and teachers with inequitable resources to the same standards, narrow and distort the curriculum, and solidify class and ethnic disparities. In addition, they argue that high-stakes testing sends the message that the primary purpose of learning is to score well on tests (Nichols & Berliner, 2008; McMillan, 2005; Mika, 2005; Burger & Krueger, 2003; FairTest, 2003; Bracey, 2000).

Whatever one's view on the value and utility of high-stakes testing, it is undeniable that these programs have had significant effects on students, teachers, and school administrators. This Information Capsule reviews research conducted on the unintended consequences of high-stakes testing.

Negative Consequences of High-Stakes Testing

Researchers have identified the following negative consequences associated with high-stakes testing programs:

- **Narrowing of the curriculum.** Studies have consistently confirmed that increasing the stakes attached to tests can change what is taught and how it is taught and adversely affect the quality of classroom practice. Studies have also found that the greater the stakes, the more likely curriculum narrowing will occur (Mesler, 2008; Moses & Nanna, 2007; Mitchell, 2006; Yeh, 2005; Sadker & Zittleman, 2004; Clarke et al., 2003; Pedulla et al., 2003; Cimbricz, 2002; Diamond & Spillane, 2002; Shepard, 2002; Stecher, 2002; American Educational Research Association, 2000; Langenfeld et al., 1997).

It should be noted that some researchers have pointed out that narrowing of the curriculum is not always undesirable. They maintain that if state standards, curriculum, and tests are well aligned, students will be taught and tested on the content and skills they are expected to know (Yeh, 2005; Steinberg, 2003; Shepard, 2002).

Researchers have identified four general categories of curriculum narrowing:

- **Exclusion of non-tested subject areas.** Studies have found that in the majority of schools, the time devoted to untested subjects such as art, foreign languages, physical education, and music has been reduced or eliminated completely so that teachers can spend more time teaching reading, math, writing, and science (Nichols & Berliner, 2008; Marchant & Paulson, 2005; Yeh,

2005; Volante, 2004; Amrein & Berliner, 2003; Westchester Institute for Human Services Research, 2003; Cimbricz, 2002; Shepard, 2002). Surveys conducted across the U.S. have found that the majority of teachers report an increase in the amount of time spent on tested subjects and a decrease in the amount of time spent on non-core subject areas and other activities, such as research projects and field trips.

- **Exclusion of non-tested topics within subject areas.** Opponents of high-stakes testing have argued that they lead to a low-level basic skills curriculum, with teachers focusing only on the parts of the curriculum that will help students earn high test scores. Studies have found that teachers often exclude topics within tested subjects that are unlikely to appear on high-stakes tests. They have also been found to change course objectives and the sequence of the curriculum to correspond to the content and timing of tests, placing more emphasis on topics that appear on the test earlier in the school year and less emphasis on topics that are not tested (Barnes, 2005; Yeh, 2005; Amrein & Berliner, 2003; Reardon & Galindo, 2002; Stecher, 2002; Bracey, 2000). Gayler and associates (2003) stated: "Many educators report that exit exams and high-stakes testing are squeezing out any content not covered by the tests, encouraging breadth of coverage instead of depth, and promoting a curriculum sequence and a pace that are not appropriate for some students."
- **Adapting teaching style to testing format.** Critics contend that high-stakes testing has caused teachers to engage in repetitious instruction on isolated pieces of information, leaving students with little time to engage in creative interdisciplinary activities or project-based inquiry. Researchers have found evidence that teachers tend to abandon more innovative instructional strategies, such as cooperative learning and creative projects, in favor of more traditional lecture and recitation to prepare students for high-stakes tests. Many teachers report adopting instructional approaches that resemble testing methods; for example, finding mistakes in written work or solving only the types of math problems contained in the test (Nichols & Berliner, 2008; Marchant & Paulson, 2005; Triplett & Barksdale, 2005; Yeh, 2005; Cimbricz, 2002; Stecher, 2002).

Clarke and colleagues' (2003) survey of teachers in three states with low-, moderate-, and high-stakes testing programs found that all teachers reported changing their instructional strategies to prepare students for high-stakes tests. However, teachers in states with higher stakes reported approximately twice the number of changes designed to adapt their instructional strategies to the test as their peers in states with low- or moderate-stakes tests.

- **Excessive test preparation.** Test preparation can benefit students by helping them build test-taking skills, such as working in isolation, listening, writing, and working within an allotted time frame. However, excessive amounts of time spent familiarizing students with the format of test questions and how to record answers results in a considerable loss of learning time. In addition, repetition of the same tasks may result in student boredom and burnout (Rhone, 2006; Yeh, 2005; Volante, 2004; Westchester Institute for Human Services Research, 2003; Cimbricz, 2002; Stecher, 2002; Ediger, 2000).

Studies indicate that teachers often devote too much classroom time to test preparation. In many states, students begin preparing for high-stakes tests when they enter school in the fall and continue until tests are administered in the spring. Only after high-stakes tests are administered do some teachers engage in the real curriculum (Triplett & Barksdale, 2005; Shepard, 2002). Pedulla and colleagues' (2003) nationwide survey of teachers found that teachers in states with high-stakes tests were more likely than their counterparts in states with lower-stakes tests to engage in test preparation earlier in the school year.

Jones and colleagues (1999) surveyed North Carolina elementary teachers and reported that 80 percent of respondents indicated that they spent over 20 percent of their instructional time practicing for End-of-Grade tests. More than 28 percent of respondents reported that they practiced for the tests over 60 percent of the time. Jones and Egley (2007) surveyed elementary teachers from 30 Florida school districts. Teachers reported that they spent an average of 43 percent of their math instructional time, 43 percent of their writing instructional time, and 38 percent of their reading instructional time on test-taking strategies specifically designed to help students score higher on the FCAT. Interestingly, there were no significant differences in the amount of time spent practicing for tests based on schools' performance grades. For example, "A" schools did not spend significantly more time practicing for tests than "C" schools. In contrast, Taylor and colleagues (2002) found that teachers at high-achieving Colorado schools reported spending less time on practice tests and test-taking strategies than teachers at lower-performing schools.

Research has demonstrated that it is possible for test scores to increase when students become familiar with the test's format, with or without real improvement in the broader achievement constructs. In these cases, students may be mistakenly categorized as high-performing not because of their actual achievement levels, but because their teachers engaged in excessive test preparation activities (Jacob, 2005; Sadker & Zittleman, 2004; Volante, 2004; Heubert, 2002).

- **Disproportionate impact on disadvantaged students.** One of the rationales given for establishing high-stakes testing programs is to improve educational equity, but some researchers have noted that these tests rarely lead to more equal educational opportunities or outcomes for students from disadvantaged backgrounds. In fact, studies have shown that the negative effects of high-stakes testing appear to be greater for low-performing, low-income, and minority students than they are for more advantaged students. This is because disadvantaged students have been found to have lower than average scores and to be more likely to live in states with strong accountability systems (Moses & Nanna, 2007; Marchant & Paulson, 2005; Nichols et al., 2005; Sadker & Zittleman, 2004; French, 2003; Gayler et al., 2003; Amrein & Berliner, 2002a; Carnoy & Loeb, 2002; Heubert, 2002; Reardon & Galindo, 2002; Stecher, 2002; Riede, 2001; Samuelson, 2001; Ediger, 2000).

High-stakes tests may reflect the inequities that exist within schools rather than meaningful differences in student learning, failing to take into account large discrepancies in school funding, facilities, parental involvement, transportation access, and neighborhood environment. States and districts must ensure that results are truly indicative of student achievement rather than a reflection of the quality of school resources or instruction. Researchers have therefore urged educators not to compare teachers and schools unless student demographics and school resources are equated (von der Embse, 2008; FairTest, 2007; Moses & Nanna, 2007; Barnes, 2005; Clarke et al., 2003; American Psychological Association, 2001; Samuelson, 2001; American Educational Research Association, 2000; National Academy of Sciences, 1999).

Studies have found that narrowing of curriculum and instruction occurs most often in low-performing, minority, and low-income schools. These schools are under the most pressure to improve test scores (Nichols & Berliner, 2008; FairTest, 2007; Volante, 2004; Shepard, 2002). Madaus and Clarke (2001) reported on surveys they administered to over 2,000 grades 4-12 teachers. They found that 75 percent of math and science teachers with high-minority classes reported pressure from their districts to improve high-stakes test scores, in comparison to approximately 60 percent of teachers with low-minority classes. Teachers with high-minority classrooms reported significantly more often that they taught test-taking skills, increased emphasis on tested topics, and began test preparation more than a month before the test. Diamond and Spillane (2002) studied high- and low-performing Chicago elementary schools. They found that high-performing schools used the results of high-stakes tests to identify and adopt interventions for all students, while low-performing schools targeted specific

students who had the greatest chance of passing the test. They concluded that high-stakes tests had a very different impact on students' educational opportunities, depending on their school's status in relation to the accountability system.

- **Misleading measure of students' overall achievement.** Major research associations have stated that high-stakes decisions should not be made on the basis of a single test score because a test only provides a "snapshot" of student achievement and may not accurately reflect an entire year's worth of student progress and achievement. They maintain that decisions about students' education, such as course placement, retention, or graduation, have more validity when they are based on additional relevant information (American Psychological Association, 2001; American Educational Research Association, 2000; National Academy of Sciences, 1999).

Most educators agree that high-stakes tests measure an extremely limited sample of knowledge and skills under the assumption that test questions are indicative of performance in broader domains. However, the number of items measuring any particular skill or knowledge may be too small to provide a reliable measure of a specific skill. In addition, students who typically perform well academically may receive a low test score because they had a bad day or undervalued the significance of the test. Researchers also point out that tests can't discern the impact of poverty, abuse, part-time jobs, lack of sleep, emotional state, or lack of social adjustment on student performance (Barnes, 2005; Marchant, 2004; Burger & Krueger, 2003; French, 2003; Heubert, 2002; Stecher, 2002).

Heubert (2002) reported that even the best standardized tests are far less precise than most people realize. He offered the following two examples:

- How often will a student who belongs at the 50th percentile according to national test norms actually score within five percentile points of that ranking on a test? The answer is only about 30 percent of the time in mathematics and 42 percent of the time in reading.
 - What are the chances that two students with identical "real achievement" will score more than ten percentile points apart on the same Stanford 9 test? For two grade 9 students who are at the 45th percentile in mathematics, the answer is 57 percent of the time. In grade 4 reading, the probability is 42 percent.
- **Test anxiety.** Test-related anxiety is one of the most commonly cited sources of student stress and it has become more prevalent as schools have attached more serious consequences to standardized testing (Ghezzi, 2010; Mesler, 2008; Wilde, 2008; Moses & Nanna, 2007; King, 2006; Rhone, 2006; Barnes, 2005; Triplett & Barksdale, 2005; New York State Department of Education, 2004; Burger & Krueger, 2003; Vogler, 2002; Dounay, 2000; Huber & Moore, 2000).

Researchers have estimated that 20 to 33 percent of all students suffer from some level of test anxiety (McCaleb-Kahan & Wenner, 2009; von der Embse, 2008). McCaleb-Kahan and Wenner (2009) reported that test anxiety is more prevalent in females than in males, regardless of grade level, and that low-income students appear to suffer from higher levels of test anxiety than more advantaged students.

Test anxiety can interfere with students' ability to function during a test and in the days and weeks leading up to a test. Physiological responses include increases in blood pressure and rate of respiration, elevated body temperature, gastrointestinal problems, headaches, difficulty sleeping, and muscle spasms. High levels of test anxiety are associated with lowered levels of academic performance, poorer study skills, and development of academic avoidance behaviors. Following testing, some overly anxious children exhibit behavioral symptoms, such as crying, illness, or outbursts of anger (Larson et al., 2010; McCaleb-Kahan & Wenner, 2009; von der Embse, 2008).

Pedulla and colleagues' (2003) national survey of teachers found that teachers from high-stakes states were more likely to report that students felt intense pressure to perform well and were extremely anxious about taking state tests. Clarke and colleagues' (2003) interviews with educators in three states (one state with low-stakes tests, one with moderate-stakes tests, and one with high-stakes tests) reported that as stakes increased, educators reported more test-related student stress. A study conducted in a California high school found that 61 percent of students reported test anxiety prior to taking the state's graduation test, with 26 percent reporting high levels of test anxiety often or most of the time (Bradley et al., 2007). Triplett and colleagues' research (Triplett & Barksdale, 2005; Triplett et al., 2003) found that elementary students worried about a variety of test-related issues, including the length of the tests, not being able to talk for long periods of time, how to bubble answer sheets, what to do when they didn't know the answers, time constraints during tests, and possible consequences of their scores. McCaleb-Kahan and Wenner (2009) reported a strong correlation between teacher anxiety and student test anxiety, leading them to conclude that teacher concerns were perceived by and transferred to students.

- **Increased pressure on teachers.** Surveys conducted across the country have found that teachers report increased pressure when high-stakes testing programs are implemented (Barnes, 2005; Clarke et al., 2003; Pedulla et al., 2003). This pressure appears to increase dramatically when high-stakes for adults, such as pay increases, job retention, or school restructuring, are attached to student test results (Barnes, 2005; Allen, 2000).

Pedulla and colleagues' (2003) nationwide survey found that teachers in high-stakes states reported feeling more pressure than those in lower-stakes states. Elementary teachers reported feeling the most test-related pressure and high school teachers reported the least pressure, with middle school teachers falling in between the two. Barnes' (2005) survey of Georgia teachers concluded that teachers' test-related pressure negatively affected their classroom performance. She noted that it is hard for teachers to focus on teaching their students when they are under the threat of losing their job, taking a pay cut, or gaining a reputation as a poor teacher.

Barksdale-Ladd and Thomas (2000) interviewed teachers in two states. Teachers reported that they felt pressure to ensure high scores on their state tests and worried about salary cuts and losing their jobs. Most interesting was the finding that although teachers ranged in experience from three to 20 years, they expressed an equal amount of frustration with the testing culture.

In Florida, Jones and Egley (2007) found that 97 percent of elementary school teachers reported feeling between "some pressure" and "a lot of pressure" to improve their students' FCAT scores. They also found that teachers who reported feeling the most pressure spent an average of 13 percent more of their instructional time teaching test-taking strategies in reading, writing, and math.

- **Lower teacher morale.** Studies have confirmed that high-stakes testing negatively affects teacher morale. Pressure to produce high test scores and threats to job security have led teachers to report a diminished sense of professional worth and feelings of disempowerment and alienation. Teachers feel that instructional decisions are increasingly based on what is most likely to be included on high-stakes tests and that they are given fewer opportunities to rely on their professional judgment and expertise (Nichols & Berliner, 2008; Perkins & Wellman, 2008; Clemmitt, 2007; Triplett & Barksdale, 2005; Volante, 2004; Burger & Krueger, 2003; Cimbricz, 2002; Stecher, 2002; Steeves et al., 2002; Langenfeld et al., 1997).

Taylor and colleagues' (2002) survey of Colorado teachers found that 81 percent of teachers reported a decrease in faculty morale that they attributed to implementation of the state's high-stakes testing program. In Ohio, a survey of National Board Certified teachers found that 80 percent of respondents believed their autonomy had declined since the introduction of high-stakes tests (Rapp, 2002). In

North Carolina, Jones and colleagues (1999) reported that 77 percent of elementary school teachers stated that their morale was lower since the introduction of high-stakes testing and over 76 percent reported that their jobs were more stressful.

- **Manipulation of student retention and reclassification policies to increase test scores.** When low-achieving students are removed from the test-taking population, the average score of the remaining students increases, even if the achievement of those actually taking the test does not improve. Unfortunately, this means that it is in some schools' best interests to eliminate lower-performing students from the testing pool. Several researchers have found that it is not uncommon for teachers to respond strategically to high-stakes testing programs by manipulating student retention and reclassification policies (Jacob, 2005; Nichols & Berliner, 2005; French, 2003; Heubert, 2002; Stecher, 2002; Steeves et al., 2002; Langenfeld, 1997).

Jacob's (2005) study of Chicago Public Schools' accountability policy found that teachers preemptively retained low-achieving students to provide them with an additional year of learning and maturation before taking high-stakes exams. Similarly, Amrein and Berliner's (2002a) study of 16 states with high school graduation exams reported that high-stakes testing policies were associated with greater numbers of low-performing students being retained to ensure that they were properly prepared to take the graduation exam. Researchers have also documented increases in the number of low-performing students being suspended or expelled before testing days. At some schools, administrators were observed reclassifying students as disabled or limited English proficient in order to remove them from the testing pool (Jacob, 2005; Amrein & Berliner, 2002b; Stecher, 2002).

Researchers have also found that some teachers at low-performing schools concentrate on the "cusp" students, or those who need only a few more points on high-stakes tests to reach proficiency levels, neglecting both the gifted and lowest-performing students in their classes (Berliner & Nichols, 2005; Booher-Jennings, 2005; Diamond & Spillane, 2002). Stecher (2002) reported that some principals reassign teachers among grade levels to improve the relative quality of teaching at high-stakes grade levels.

Finally, states can reduce failure rates by setting lower cut-off scores for proficiency or by making graduation tests easier to pass (Heubert, 2002). Berliner and Nichols (2005) stated: "The greater the pressure to perform at a certain level, the more likely people will find a way to distort and corrupt the system to achieve favorable results."

Positive Consequences of High-Stakes Testing

So far, the unintended consequences of high-stakes testing discussed in this report have all been negative. Researchers do agree, however, that high-stakes testing programs have had some positive effects on schools and districts.

- **Increased teacher professional development.** There is some evidence that high-stakes testing has led to more focused teacher professional development (McMillan, 2005; Stecher, 2002; Cizek, 2001; Allen, 2000). Pedulla and colleagues' (2003) survey of teachers nationwide found that greater proportions of teachers in states where testing stakes were low indicated that there was no professional development related to test preparation, interpretation, and use of results. Jones and Egley's (2007) online survey of elementary teachers from 30 school districts in Florida found that teachers with more professional development support were more likely than teachers with less support to believe that the FCAT had a positive impact on their ability to use effective teaching methods. In contrast to these findings, however, Taylor and colleagues' (2002) survey of Colorado teachers reported that only 32 percent agreed that professional development opportunities increased with the introduction of the state's high-stakes testing program.

Cizek (2001) suggested that high-stakes testing programs have created a corresponding increase in teachers' knowledge of testing. He noted: "Increasingly, teachers can tell you the difference between a norm-referenced test and a criterion-referenced test; they can recognize, use, or develop a high-quality rubric; they can tell you how their state's writing test is scored, and so on."

- **Alignment of instruction with state content standards.** Studies have found that high-stakes testing leads teachers to better align instruction with state content standards to ensure that students are taught and tested on the content and skills they are expected to master (Perkins & Wellman, 2008; Yeh, 2005; Gayler et al., 2003; Stecher, 2002). Pedulla and colleagues' (2003) nationwide survey found that the vast majority of teachers indicated that their district's curriculum was aligned with state tests. Taylor and colleagues' survey found that Colorado teachers reported gradually aligning curricula with state standards. In Wise and colleagues' (2003) survey of teachers and administrators in California schools, staff reported that high school instruction covered an increasingly greater portion of the content standards assessed by the exit exam. In 1999, only about 20 percent of schools reported covering at least 75 percent of the standards; by 2003, over 80 percent of the schools reported at least 75 percent coverage. A number of middle and senior high schools also reported introducing new courses and adopting new textbooks for existing courses in order to better align their instruction with state content standards.
- **More opportunities for remediation.** High-stakes tests provide schools with the opportunity to diagnose student weaknesses and develop the instructional programs needed to help low-achieving students meet state standards on subsequent tests. Researchers have noted that many schools combine high-stakes testing programs with a meaningful system of supports, including remediation and prevention services for low-achieving students (Barnes, 2005; Burger & Krueger, 2003; Gayler et al., 2003; American Psychological Association, 2001; Rabinowitz et al., 2001; American Educational Research Association, 2000; Lewis, 2000; Walker, 2000). Wise and colleagues' (2003) survey on the impact of California's high school exit exam found that teachers and administrators reported adding a number of new remedial or supplemental courses, including many aimed specifically at students who did not pass the exit exam on their first attempt, as well as limited English proficient and disabled students.
- **Increased use of data to inform instruction.** Advocates of high-stakes testing point out that test results give classroom teachers important information on how well individual students are learning. Pedulla and colleagues' (2003) nationwide survey found that teachers reported using test results to the greatest extent when the stakes for statewide testing programs were high. For example, significantly more teachers (40 percent) in states with high-stakes for schools and students than in low-stakes states (10 percent) reported that their schools' results influenced their teaching on a daily basis. More teachers in states with high-stakes tests reported using the results to plan instruction (60 percent) and to select instructional materials (50 percent) than teachers in low-stakes states (40 and 30 percent, respectively).

Jones and Egley's (2007) online survey of elementary school teachers from 30 Florida school districts asked teachers "How useful are the FCAT results for helping you assess students' strengths and weaknesses?" Teachers did not find the results to be particularly useful in reading, writing, or mathematics. For example, in writing, over 22 percent of respondents agreed that the results were "not useful at all" and only 2.6 percent agreed that the results were "very useful." In math, over 18 percent of teachers agreed that the results were "not useful at all" and only 2.7 percent agreed that the results were "very useful."

Studies suggest that the time of year in which test score reports are distributed to teachers is important. Clarke and colleagues' (2003) interviews with educators in three states found that approximately 20 percent of interviewees in each state believed that high-stakes test results came back too late to be

useful. McMillan's (2005) survey of school districts in Virginia found that when results were given to teachers at the end of the school year, they didn't have time to use the results to make changes in instruction. Similarly, distribution of results at the beginning of the new school year was not effective because teachers were busy with opening of school activities. He found some evidence that providing teachers with their future students' score reports in mid to late summer, when they had time to study the results and think about the implications, was related to more positive changes in instruction.

Inconsistent Findings

As summarized below, the research is mixed on the impact of high-stakes testing on dropout rates, students' levels of academic achievement and motivation, and on the consequences of publishing test scores.

- **Dropout rates.** Researchers have not reached consensus as to whether high school graduation tests lead to higher dropout rates. Some studies have concluded that dropout rates increase when students are required to pass a test in order to graduate from high school. These studies suggest that the increased test preparation and narrower curriculum causes low-achieving students to become more disillusioned with the educational process and increase their disengagement from school. Other students try as hard as they can, but still cannot pass the exit exam (Nichols & Berliner, 2008; Marchant & Paulson, 2005; Amrein & Berliner, 2002a; Rabinowitz et al., 2001; Samuelson, 2001).

On the other hand, several studies have found that administration of high school exit exams has no impact on dropout rates. Their findings suggest that low-achieving students disengage from school well before they are required to take graduation tests and that high school exit exams do not influence their decision to drop out of school (American Educational Research Association, 2009; New York State Education Department, 2004; Westchester Institute for Human Services Research, 2003; Heubert, 2002; Rabinowitz et al., 2001; Langenfeld et al., 1997). In fact, Davenport and colleagues (2002) found that over half of Minnesota students who dropped out had already passed both portions of the graduation test, leading the authors to conclude that for a substantial number of dropouts, passing the exit exams was not the determining factor in their decision to leave school. Greene and Winters (2004) reported: "In most states students are routinely given second, third and even seventh chances to pass exit exams before they are finally denied a diploma. Between each administration of the test, students who have failed are provided with extra help specifically designed to get them past the test requirement. Given so many tries, eventually most students who are able to complete the other requirements to graduate also pass the exit exam, even if only by chance."

Some studies indicate that high-stakes testing has different effects on students, depending on their academic achievement levels, ethnicity, and socioeconomic status. For example, Haney (in Clarke et al., 2000) reported that dropout rates increased the year Texas' standards-based exit exam was introduced, but that decreases in high school completion rates were approximately 50 percent greater for Black and Hispanic students than for White students. Reardon and Galindo (2002) found that high-stakes tests were associated with a larger increase in the probability of dropping out for low-income and low-achieving students nationwide. Clarke, Haney, and Madaus (2000) found that failing Florida's high school graduation test was only associated with a significant increase in the likelihood of dropping out for students with moderately good grades (in the range of 1.5 to 2.5 on a 4-point scale).

Haney's examination of Texas dropout patterns over a 20 year period (summarized in Clarke et al., 2000) suggests that dropout rates may be influenced by the type of exit exam administered (minimum competency versus more rigorous standards-based tests). During this time, the Texas Educational Assessment of Minimum Skills (TEAMS), a minimum competency test, was replaced by the more difficult Texas Assessment of Academic Skills (TAAS). Haney found that introduction of the TEAMS

did not dramatically affect high school completion rates; however, dropout rates increased the year the TAAS was introduced as a requirement for high school graduation.

In any case, it has been very difficult to prove a causal connection between high-stakes tests and dropout rates. Existing studies are correlational in nature, confirming only that there is a relationship between the two variables, not that exit exams actually cause higher dropout rates. Furthermore, the types of exit exams administered vary widely from state to state and there is no uniformity among states in how dropouts are defined and counted, making it difficult to compare the results obtained in different states.

Gayler and associates (2003) noted that policies such as retaining students in grade and instituting tougher course requirements may be more strongly associated with dropping out of school than exit exams. They concluded, therefore, that while exit exams may not be the primary reason students drop out, they may be a contributing factor for some. The New York State Education Department (2004) concluded: "Decisions to drop out of school are probably very dynamic, complex processes and therefore not simple ones to model or simulate."

- **Academic achievement.** Studies conducted on the impact of high-stakes testing on student achievement have reported inconsistent findings. Some researchers have found that higher scores on high-stakes tests do not translate into higher scores on other standardized tests. They argue that students learn the content of high-stakes tests and little else (Nichols & Berliner, 2008; McMillan, 2005; Amrein & Berliner, 2002a; Heubert, 2002; Shepard, 2002; Wright, 2002). However, other researchers have found that high-stakes tests lead to substantial increases in students' scores on independent standardized tests such as the National Assessment of Educational Progress and the Iowa Test of Basic Skills (Jacob, 2005; Carnoy & Loeb, 2003; Raymond & Hanushek, 2003; Rosenshine, 2003). The New York State Education Department (2004) hypothesized that high-stakes tests are often associated with higher scores on other tests because they are usually part of a larger package of educational reforms that are implemented simultaneously and generally include establishing content standards, changes in instruction, and consequences for teachers and/or schools.

One study of special interest was conducted by Winters, Greene, and Trivitt (2008). The researchers used a data set provided by the Florida Department of Education to analyze the FCAT scores of fifth grade Florida public schools students. They found that the reading and math gains of students enrolled in F-graded schools exceeded those of students enrolled in D-graded schools, but no significant differences were found in the gains of A, B, C, and D-graded schools. Winters and colleagues suggested that the F grade and its stronger sanctions had a positive impact on student performance. Although science was not tested on the FCAT at the time of this study, the researchers found that F-graded schools also posted greater test score gains in science. They hypothesized that accountability testing led schools to adopt reforms that improved their overall quality of instruction in both high-stakes and low-stakes subject areas. The reader is cautioned that this study supports only the existence of a relationship between school performance grades and test score gains, but does not provide evidence that schools' receipt of an F-grade actually caused greater test score gains. For example, greater gains by students attending F-graded schools could have been due to regression to the mean, the statistical phenomenon that students who score at the extremes on a test are likely to score closer to the mean the second time the test is administered. In this case, students who received the lowest scores would have been more likely to receive higher scores the next time they were tested.

- **Student motivation.** Critics of high-stakes tests maintain that they damage students' intrinsic motivation and encourage them to focus on test performance instead of the learning process. They contend that an emphasis on learning for the sake of succeeding on tests discourages students from exploring subjects that interest them and communicates to students that their other abilities (such as foreign languages, history, dance, or computer programming) are of little or no value (Mesler, 2008; Nichols &

Berliner, 2008; Clemmitt, 2007; Marchant, 2004; Volante, 2004; Clarke et al., 2003; Stecher, 2002; Madaus & Clarke, 2001).

Researchers have documented that many students who fear they won't receive high scores give up too easily, while those who are confident in their ability to earn high scores often lose interest in school because they are not intellectually challenged by the test-based curriculum. Some students report not being motivated to work toward test success because they don't see a connection between high test scores and college or job prospects (Clemmitt, 2007; New York State Education Department, 2004; Volante, 2004; Amrein & Berliner, 2003; FairTest, 2003; Madaus & Clarke, 2001).

In contrast, advocates maintain that high-stakes tests motivate all students to study and work harder in order to earn high test scores. They argue that attaching consequences to tests encourages students to take their coursework more seriously (Elbousty, 2009; Perkins & Wellman, 2008; Barnes, 2005; Marchant & Paulson, 2005; New York State Education Department, 2004; Stecher, 2002; Rabinowitz et al., 2001; Walker, 2000). More research is clearly needed to determine how high-stakes testing affects student motivation.

- **Information provided to the public.** Researchers contend that the attention given to high-stakes tests by the media, state departments of education, and school district personnel serves to elevate them into even higher-stakes status. Supporters of high-stakes tests argue that they provide the accountability that the public demands, allowing the community to evaluate how well schools are performing and identify disparities among schools. They point out that test scores can be used as a basis for implementing changes that will increase the public's confidence in schools (Burger & Krueger, 2003; Riede, 2001; Samuelson, 2001; American Educational Research Association, 2000; Walker, 2000). Rabinowitz, Zimmerman, and Sherman (2001) noted: "Assessment is the most visible and quantifiable piece of the standards-and-accountability reform. When schools begin designing curricula based on new state standards, parents, policymakers, and other members of the public may or may not take note. But when large numbers of students start failing the new tests, people begin paying attention, especially parents and the news media."

Opponents of high-stakes testing, on the other hand, contend that the publication of scores leads the public to adopt a simplistic view of education. They argue that the attention paid to test scores sends the message that they are the only important indicator of school performance, when in actuality they represent only a small piece of what the public needs to know about schools. Furthermore, critics maintain that an emphasis on test scores shifts the public's attention from the teaching and testing of the knowledge and skills that are most important for students to master to those that are easiest to measure (FairTest, 2007; Volante, 2004; Amrein & Berliner, 2003; Cimbricz, 2002; Stecher, 2002; Steeves et al., 2002).

Summary

High-stakes testing is one of the most controversial issues in American education. While some policymakers view testing as an important component of the school reform and improvement process, others see it as a threat to the quality of teaching and learning. Whatever one's view on the value and utility of high-stakes testing, it is undeniable that these programs have had significant effects on students, teachers, and school administrators. This Information Capsule summarized research conducted on the unintended consequences of high-stakes testing. Studies have found that high-stakes testing programs have a disproportionate impact on disadvantaged students and are often a misleading measure of students' overall achievement. In addition, they can lead to the following negative consequences: narrowing of the curriculum; test anxiety; increased pressure on teachers; lower teacher morale; and manipulation of retention and reclassification policies to increase scores.

Studies have found that high-stakes testing programs have positive effects on schools and districts, including increased teacher professional development; better alignment of instruction with state content standards; more remediation for low-achieving students; and increased use of data to inform instruction.

Research is mixed on the impact of high-stakes testing on dropout rates, students' levels of academic achievement and motivation, and on the consequences of publishing test scores. More research is clearly needed in all of these areas before any definitive conclusions can be drawn.

- Studies have been unable to determine whether exit exams lead to higher dropout rates. Some studies have concluded that dropout rates increase when students are required to pass a test in order to graduate from high school, while others have found that administration of high school exit exams has no impact on dropout rates.
- Some researchers have found that score increases on high-stakes tests do not transfer over or generalize to other tests, leading them to argue that students learn the content of high-stakes tests and little else. Other researchers have found that high-stakes tests lead to substantial increases in students' scores on other standardized tests.
- Some studies have found that high-stakes tests damage students' intrinsic motivation and discourage them from exploring subjects that are not included on the tests. Others have concluded that high-stakes tests motivate students to study and work hard.
- Researchers agree that publication of high-stakes test scores allows the community to evaluate how well schools are performing and identify disparities among schools. However, educators should be aware that the publication of scores may send the message that test scores are the only important indicator of school performance and lead the public to accept an over-simplified view of education.

A brief review of studies examining the financial costs of high-stakes testing programs is presented in the following section. Researchers have concluded that the direct and indirect costs of high-stakes testing are substantial and that these tests are not a low-cost solution to educational problems. One study noted that the resources needed to improve student achievement, such as instructional time and staff time, have been diverted away from teaching and learning and reinvested in test preparation, administration, and reporting. Studies have found that the overwhelming majority of high-stakes testing costs are borne at the local school district level. Although states traditionally finance expenses such as test development, shipping and return of materials, and scoring, districts are responsible for significant costs associated with test preparation and administration. The largest share of district-level testing costs are related to remediation of failing students, expenditures to prevent school failure, and teacher professional development.

How Much Does It Cost to Implement a High-Stakes Testing Program?

Before the No Child Left Behind (NCLB) Act was signed into law in January 2002, states collectively spent less than \$423 million on standardized tests. This figure increased dramatically after the passage of NCLB. By the 2007-08 school year, states were estimated to spend almost \$1.1 billion on these tests (Vu, 2008). Few studies, however, have thoroughly examined the costs of high-stakes test development and implementation.

The Florida Department of Education (2010) estimated that during the 2009-10 school year, administration of all components of the FCAT cost \$29.41 per student. This figure included development of test questions, holding review meetings with Florida educators, field testing, production and printing of testing, shipping and return of materials, scoring, reporting scores to parents, schools, and districts, and analysis and research. Although \$29.41 per student does not sound like a lot of money, when multiplied by over 2.5 million Florida public school students, the state spends an estimated \$74 million each year to test students on the FCAT. This estimate does not take into account any of the associated testing costs that are assumed by local school districts, including student remediation, teacher professional development, and loss of instructional and personnel time.

Studies have found that the overwhelming majority of high-stakes testing costs are borne at the local level. Testing costs for which local school districts are responsible include (New York State Education Department, 2004; Gayler et al., 2003; Rose & Myers, 2003):

- **Prevention:** Expenditures to prevent school failure, such as revamping instruction; developing instructional techniques to present standards-based material to special education, limited English proficient, and at-risk students; instituting early learning programs; and tracking individual students' skill levels.
- **Professional development:** Training teachers to teach in a standards-based environment; conduct test administrations; and use test scores to diagnose student weaknesses and revise instruction.
- **Remediation:** The cost of programs for students who failed the test, including summer school; after-school programs; tutoring; development of a remedial curriculum focused on core skills; remedial software; and other needed academic intervention services.
- **Testing/Administration:** Expenses to develop and disseminate information about tests; administer tests; keep records and analyze student strengths and weaknesses; create alternative exams; provide accommodations for students with disabilities; and arrange for retests.

The Center on Education Policy commissioned studies to estimate the state and local costs of implementing high school exit exams (Gayler & Kober, 2004; Gayler et al., 2003; Rose & Myers, 2003). Expert panels concluded that the costs of exit exams in Indiana, Massachusetts, and Minnesota were substantial, although they varied among states.

- In Massachusetts, it cost an estimated \$385 per student per year to implement the exit exam component of the Massachusetts Comprehensive Assessment System.
- In Minnesota, it cost \$171 per student per year to implement the Minnesota Basic Skills Test.
- In Indiana, it cost \$557 per student per year to implement the Graduation Qualifying Exam.

- State variations reflected differences in the difficulty of the exams, their initial student pass rates, and their degree of connection to the state’s broader reforms. The researchers concluded that in an average state, the cost difference between easier exams (exams geared to a sixth through eighth grade level) and harder exams (exams aligned to tenth grade standards) was about \$280 per student per year.
- Center on Education Policy researchers concluded that states should stop treating exit exams as if they are a low-cost or no-cost solution to educational problems.

The Center on Education Policy studies found that local school districts were responsible for the vast majority of costs associated with exit exams.

- In all three states, more than 96 percent of the costs were borne at the local level. The state level costs associated with exit exams were approximately \$2 per student in Indiana and Minnesota and \$7 in Massachusetts. This means that in Indiana, for example, local school districts assumed \$555 per student per year in exit exam costs, while the state’s exam costs averaged \$2 per student per year.
- In all three states studied, the direct testing costs of developing and administering exit exams constituted less than 20 percent of the exam costs (less than 7 percent in Minnesota, 9 percent in Massachusetts, and 18 percent in Indiana). The bulk of the costs went toward remediation, prevention, and professional development activities to help students pass the exams (94 percent in Minnesota, 91 percent in Massachusetts, and 82 percent in Indiana).

The Center on Education Policy studies concluded that the costs of helping English language learners (ELL) and students with disabilities pass exit exams were significant. In Indiana, recalculation of exit exam costs to include students with disabilities and ELLs raised the estimates by \$135 per student per year (from \$422 to \$557). In Massachusetts, ELLs added about \$101 per student per year to the total costs associated with the exit exam for both ELL and non-ELL students. In Minnesota, the exit exam was estimated to cost an extra \$26 per year for each special education student. Since Minnesota’s exit exam focused on eighth grade skills, researchers hypothesized that the costs of preparing special education students in states with exams aligned to higher standards would be much greater.

The Center on Education Policy studies also found that costs rose when states increased passing rates on their current exit exams, raised the cut score for proficient performance, or switched to a more challenging exam.

- In Minnesota, the cost of moving from measurement of eighth to tenth grade standards, while maintaining the established passing rate, was estimated at an additional \$377 per student.
- In Massachusetts, the cost of raising the minimum score required for graduation from the “needs improvement level” to the “proficient level” was estimated to add exam-related expenses of \$575 per student per year.
- In Indiana, the cost to raise the passing rate to the state standard of “commendable performance” was estimated to add \$685 per student per year to the cost of the exam.
- The researchers found that it was actually less expensive to start with a rigorous exit exam than to change from an easier to a harder test at some future point.

Zellmer, Frontier, and Pheifer (2006) estimated that Wisconsin's school districts spent an average of \$33.91 per student on the Wisconsin Knowledge and Concepts Exam (WKCE), extrapolated to \$14,700,000 for WKCE testing statewide in 2005-06. This figure did not include money spent by the Wisconsin state education agency for development, publication, shipment to and from schools, scoring services, and reporting of student results. The researchers also surveyed administrators from 171 Wisconsin school districts to determine the costs of testing in terms of time and human resources and categorized their findings into the following three areas:

- **Logistical Preparation:** Staff spent a per-district average of over 91 hours placing testing materials in secure locations; verifying and preparing labels and affixing them to test booklets; distributing and collecting testing materials; packing and shipping test booklets and answer sheets for scoring; and preparing schedules and managing logistics.
- **Test administration:** Paraprofessionals spent a per-district average of 102 hours assisting teachers with testing. Teachers spent a per-district average of 976 hours administering the tests. Administrators spent a per-district average of 62 hours engaged in test-related tasks. Across all districts in the sample, over one thousand substitute teachers were hired to proctor tests or supervise the classrooms of teachers engaged in other testing activities.
- **Loss of instructional time:** Districts reported between six and nine days of disrupted instructional services. Some schools reported that disadvantaged student populations (special education, Title I, and limited English proficient students) lost as much as 15 days (three weeks) of instructional time.

Zellmer and colleagues (2006) concluded that the resources needed to improve student achievement, such as instructional time and staff time, have been diverted away from teaching and learning and reinvested in test preparation, administration, and reporting. Marchant (2004) noted: "Costs come not only on a financial level but also at a personal level for the children, and at a professional level for educators. At the financial level, the costs are high."

All reports distributed by Research Services can be accessed at <http://drs.dadeschools.net>.

References

- AbdAllen, M. (2000). High-Stakes Testing Pressures Teacher Education. *State Education Leader*, 18(1), 16-20. ERIC Document Reproduction Service No. ED452219.
- American Educational Research Association. (2000). *AERA Position Statement on High-Stakes Testing in Pre-K - 12 Education*. Retrieved from <http://www.aera.net/?id=378>.
- American Psychological Association. (2001). *Appropriate Use of High-Stakes Testing in Our Nation's Schools*. Retrieved from <http://www.apa.org/pubs/info/brochures/testing.aspx?item=1>.
- Amrein, A.L., & Berliner, D.C. (2002a). *The Impact of High-Stakes Tests on Student Academic Performance: An Analysis of NAEP Results in States With High-Stakes Tests and ACT, SAT, and AP Results in States With High School Graduation Exams*. Education Policy Research Unit, Arizona State University, Tempe, AZ. Retrieved from <http://www.epicpolicy.org/files/EPSSL-0211-126-EPRU.pdf>.
- Amrein, A.L., & Berliner, D.C. (2002b). *An Analysis of Some Unintended and Negative Consequences of High-Stakes Testing*. Education Policy Research Unit, Arizona State University, Tempe, AZ. Retrieved from <http://epsl.asu.edu/epru/documents/EPSSL-0211-125-EPRU-exec.pdf>.
- Amrein, A.L., & Berliner, D.C. (2003). The Effects of High-Stakes Testing on Student Motivation and Learning. *Educational Leadership*, 60(5), 32-38.
- Barksdale-Ladd, M.A., & Thomas, K.F. (2000). What's At Stake in High-Stakes Testing: Teachers and Parents Speak Out. *Journal of Teacher Education*, 51(5), 384-397.
- Barnes, M. (2005). The Discriminatory Effects of High-Stakes Testing in Georgia: Exploring Causes and Solutions. *Education Law and Policy Forum*, Volume 1. Education Law Consortium, The University of Georgia, Athens, GA. Retrieved from <http://www.educationlawconsortium.org/forum/2005/papers/barnes.pdf>.
- Berliner, D.C., & Nichols, S.L. (2005). *Test Results Untrustworthy*. Education Policy Research Unit, Arizona State University, Tempe, AZ. ERIC Document Reproduction Service No. ED508512.
- Booher-Jennings, J. (2005). Below the Bubble: "Educational Triage" and the Texas Accountability System. *American Educational Research Journal*, 43(2), 231-268.
- Bracey, G. (2000). *High Stakes Testing*. Center for Education Research, Analysis, and Innovation, School of Education, University of Wisconsin-Milwaukee, Milwaukee, WI. Retrieved from <http://nepc.colorado.edu/files/cerai-00-32.htm>.
- Bradley, R.T., McCraty, R., Atkinson, M., Arguelles, L., Rees, R.A., & Tomasino, D. (2007). *Reducing Test Anxiety and Improving Test Performance in America's Schools: Summary of Results from the TestEdge National Demonstration Study*. Institute of HeartMath, Boulder Creek, CA. Retrieved from http://www.betaheartmath.org/templates/ihtm/section_includes/education/pdf/Tends_Summary_Results.pdf.
- Burger, J.M., & Krueger, M. (2003). A Balanced Approach to High-Stakes Achievement Testing: An Analysis of the Literature With Policy Implications. *International Electronic Journal for Leadership in Learning*, 7(4). Retrieved from http://www.ucalgary.ca/iejll/burger_krueger/.
- Carnoy, M., & Loeb, S. (2002). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Educational Evaluation and Policy Analysis*, 24(4), 305-331.

- Cimbricz, S. (2002). State-Mandated Testing and Teachers' Beliefs and Practice. *Education Policy Analysis Archives*, 10(2). Retrieved from <http://epaa.asu.edu/epaa/v10n2.html>.
- Cizek, G.J. (2001). More Unintended Consequences of High-Stakes Testing. *Educational Measurement: Issues and Practice*, 20(4), 19-27.
- Clarke, M., Haney, W., & Madaus, G. (2000). *High Stakes Testing and High School Completion*. National Board on Educational Testing and Public Policy, Lynch School of Education, Boston College, Chestnut Hill, MA. Retrieved from <http://www.bc.edu/research/nbetpp/publications/v1n3.html>.
- Clarke, M., Shore, A., Rhoades, K., Abrams, L., Miao, J., & Li, J. (2003). *Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from Interviews with Educators in Low-, Medium-, and High-Stakes States*. National Board on Educational Testing and Public Policy, Lynch School of Education, Boston College, Chestnut Hill, MA. Retrieved from <http://www.bc.edu/research/nbetpp/statements/nbr1summary.pdf>.
- Clemmitt, M. (2007). Students Under Stress. *CQ (Congressional Quarterly) Researcher*, 17(25), 577-600.
- Davenport, E.C., Davison, M.L., Kwak, N., Irish, M.L., & Chan, C-K. (2002). *Minnesota High Stakes High School Graduation Test and Completion Status for the Class of 2000*. Office of Educational Accountability, College of Education and Human Development, University of Minnesota, Minneapolis, MN.
- Diamond, J.B., & Spillane, J.P. (2002). *High Stakes Accountability in Urban Elementary Schools: Challenging or Reproducing Inequality?* Institute for Policy Research Working Paper, Northwestern University, Evanston, IL. Retrieved from <http://www.northwestern.edu/ipr/publications/papers/2002/WP-02-22.pdf>.
- Dounay, J. (2000). High-Stakes Assessments Bring Out the Critics. *State Education Leader*, 18(1), 4-6. ERIC Document Reproduction Service No. ED452219.
- Ediger, M. (2000). *High Stakes Testing and Student Achievement*. ERIC Document Reproduction Service No. ED447193.
- Elbousty, Y. (2009). *High Stakes Testing Literature Review and Critique*. Paper presented at the Annual Meeting of the Northeastern Educational Research Association, Rocky Hill, CT, October 2009. Retrieved from http://digitalcommons.uconn.edu/cgi/viewcontent.cgi?article=1013&context=nera_2009.
- FairTest. (2003). *Reports: High Stakes Testing Hurts Education*. Retrieved from <http://www.fairtest.org/reports-high-stakes-testing-hurts-education>.
- FairTest. (2007). *The Dangerous Consequences of High-Stakes Standardized Testing*. Retrieved from <http://www.fairtest.org/dangerous-consequences-highstakes-standardized-tes>.
- Florida Department of Education. (2010). *Frequently Asked Questions: Assessment and School Performance*. Retrieved from <http://www.fldoe.org/faq/default.asp?Dept=179&ID=977>.
- French, D. (2003). A New Vision of Authentic Assessment to Overcome the Flaws in High Stakes Testing. *Middle School Journal*, 35(1). Retrieved from <http://www.cce.org/french.ms1.9.03.doc>.
- Gayler, K., & Kober, N. (2004). The Hidden Costs of High School Exit Exams. In *Pay Now or Pay Later: The Hidden Costs of High School Exit Exams*. Center on Education Policy, Washington, DC. Retrieved from http://www.edlawcenter.org/ELCPublic/elcnews_090611_HiddenCosts.pdf.

- Gayler, K., Chudowsky, N., Kober, N., & Hamilton, M. (2003). *State High School Exit Exams Put to the Test*. Center on Education Policy, Washington, DC. Retrieved from http://www.schoolfunding.info/resource_center/research/CEPExitExamStudy.pdf.
- Ghezzi, P. (2010). *Help Your Child Reduce Test Stress*. Retrieved from <http://www.schoolfamily.com/school-family-articles/article/10697-help-your-child-reduce-test-stress>.
- Greene, J.P., & Winters, M.A. (2004). The Positive Effects of Exit Exams. *The Indianapolis Star*, May 2, 2004.
- Heubert, J.P. (2002). *High-Stakes Testing: Opportunities and Risks*. National Center on Accessing the General Curriculum, Wakefield, MA. Retrieved from http://www.cast.org/publications/ncac/ncac_highstakes.html.
- Huber, R.A., & Moore, C.J. (2000). Education Reform Through High Stakes Testing - Don't Go There. *Science Educator*, 9(1), 7-13.
- Jacob, B.A. (2005). Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools. *Journal of Public Economics*, 89(5-6), 761-796.
- Jones, B.D., & Egley, R.J. (2007). Learning to Take Tests or Learning for Understanding? Teachers' Beliefs about Test-Based Accountability. *Educational Forum*, 71(3), 232-248.
- Jones, M.G., Jones, B.D., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The Impact of High-Stakes Testing on Teachers and Students in North Carolina. *Phi Delta Kappan*, 81(3), 199-203.
- King, L. (2006). Schools Reduce Test Stress Through Relaxation Techniques. *USA Today*, March 7, 2006.
- Langenfeld, K., Thurlow, M., & Scott, D. (1997). *High Stakes Testing for Students: Unanswered Questions and Implications for Students with Disabilities*. National Center on Educational Outcomes, University of Minnesota, Minneapolis, MN. ERIC Document Reproduction Service No. ED415627.
- Larson, H.A., El Ramahi, M.K., Conn, S.R., Estes, L.A., & Ghibellini, A.B. (2010). Reducing Test Anxiety Among Third Grade Students Through the Implementation of Relaxation Techniques. *Journal of School Counseling*, 8(19). ERIC Document Reproduction Service No. EJ885222.
- Lewis, A. (2000). High-Stakes Testing: Trends and Issues. *Mid-Continent Research for Education and Learning Policy Brief*, April 2000. Retrieved from http://www.mcrel.org/pdf/policybriefs/5993pi_pbhighstakestesting.pdf.
- McCaleb-Kahan, P., & Wenner, R. (2009). *The Relationship of Student Demographics to 10th Grade MCAS Test Anxiety*. Paper presented at the 40th Annual Meeting of the Northeastern Educational Research Association, Rocky Hill, CT, October 2009. Retrieved from http://digitalcommons.uconn.edu/cgi/viewcontent.cgi?article=1002&context=nera_2009.
- McMillan, J.H. (2005). *The Impact of High-Stakes Test Results on Teachers' Instructional and Classroom Assessment Practices*. Metropolitan Educational Research Consortium, Virginia Commonwealth University, Richmond, VA. ERIC Document Reproduction Service No. ED490648.
- Madaus, G., & Clarke, M. (2001). *The Adverse Impact of High Stakes Testing on Minority Students: Evidence from 100 Years of Test Data*. ERIC Document Reproduction Service No. ED450183.
- Marchant, G.J. (2004). What Is At Stake With High Stakes Testing? A Discussion of Issues and Research. *The Ohio Journal of Science*, 104(2), 2-7.

- Marchant, G.J., & Paulson, S.E. (2005). The Relationship of High School Graduation Exams to Graduation Rates and SAT Scores. *Education Policy Analysis Archives*, 13(6). ERIC Document Reproduction Service No. EJ846516.
- Mesler, A. (2008). *High Stakes and Higher Impacts: The Effects of High-Stakes Testing on School and Classroom Operations*. Duke University, Durham, NC. Retrieved from http://uwp.duke.edu/uploads/assets/Mesler_09.pdf.
- Mika, S. (2005). *Testing Testing: Do Schools Do Too Much?* Gallup, Inc. Retrieved from <http://www.gallup.com/poll/19552/testing-testing-schools-too-much.aspx>.
- Mitchell, R. (2006). *Research Review: Effects of High-Stakes Testing on Instruction*. The Center for Public Education, Alexandria, VA. Retrieved from <http://www.centerforpubliceducation.org>.
- Moses, M.S., & Nanna, M.J. (2007). The Testing Culture and the Persistence of High Stakes Testing Reforms. *Education and Culture*, 23(1), 55-72.
- Mulvenon, S.W., Connors, J.V., & Lenares, D. (2001). *Impact of Accountability and School Testing on Students: Is There Evidence of Anxiety?* Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Little Rock, AR, November 2001. ERIC Document Reproduction Service No. 460155.
- National Academy of Sciences. (1999). *High Stakes Testing and Social Promotion*. Statement of Robert M. Hauser, Ph.D. before the Committee on Health, Education, Labor, and Pensions, United States Senate. Retrieved from http://www7.nationalacademies.org/ocga/testimony/Elem_Sec_Education_Act_Reauth.asp.
- New York State Education Department. (2004). The Impact of High-Stakes Exams on Students and Teachers. *Policy Brief*, August 2004. Retrieved from <http://www.oms.nysed.gov/faru/TheImpactofHighStakesExams.htm>.
- Nichols, S. (2003). *Motivation for High Stakes Testing*. University of Michigan, Ann Arbor, MI. Retrieved from <http://sitemaker.umich.edu/356.nichols/home>.
- Nichols, S.L., & Berliner, D.C. (2005). *The Inevitable Corruption of Indicators and Educators Through High-Stakes Testing*. Education Policy Research Unit, Arizona State University, Tempe, AZ. Retrieved from <http://epicpolicy.org/files/EPSSL-0503-101-EPRU.pdf>.
- Nichols, S.L., & Berliner, D.C. (2008). Testing the Joy Out of Learning. *Educational Leadership*, 65(6), 14-18.
- Nichols, S.L., Glass, G.V., & Berliner, D.C. (2005). *High-Stakes Testing and Student Achievement: Problems for the No Child Left Behind Act*. Education Policy Research Unit, Arizona State University, Tempe, AZ. Retrieved from <http://epicpolicy.org/files/EPSSL-0509-105-EPRU.pdf>.
- Pedulla, J.J., Abrams, L.M., Madaus, G.F., Russell, M.K., Ramos, M.A., & Miao, J. (2003). *Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers*. National Board on Educational Testing and Public Policy, Lynch School of Education, Boston College, Chestnut Hill, MA. Retrieved from <http://www.bc.edu/research/nbetpp/statements/nbr2summary.pdf>.
- Perkins, G., & Wellman, N. (2008). The Impact of High-Stakes Testing: Recommendations for Professional School Counselors and Other Support Personnel. *Academic Leadership*, 8(3). Retrieved from http://www.academicleadership.org/emprical_research/505.shtml.

- Rabinowitz, S., Zimmerman, J., & Sherman, K. (2001). *Do High Stakes Tests Drive Up Student Dropout Rates? Myths Versus Reality*. WestEd Knowledge Brief, San Francisco, CA. ERIC Document Reproduction Service No. ED463565.
- Rapp, D. (2002). National Board Certified Teachers In Ohio Give State Education Policy, Classroom Climate, and High-Stakes Testing a Grade of F. *Phi Delta Kappan*, 84(3), 215-218.
- Raymond, M.E., & Hanushek, E.A.. (2003). High-Stakes Research. *Education Next*, 3(3), 48-55.
- Reardon, S.F., & Galindo, C. (2002). *Do High-Stakes Tests Affect Students' Decisions to Drop Out of School? Evidence from NELS*. Population Research Institute, University Park, PA. ERIC Document Reproduction Service No. ED482665.
- Rhone, A.E. (2006). Preparing Minority Students for High-Stakes Tests: Who Are We Cheating? *Childhood Education*, 82(4), 233-235.
- Riede, P. (2001). Testing Dissidents: School Leaders Go Public With Their Concerns Over the Harm of Highstakes Tests. *The School Administrator*, 58(11), December 2001.
- Rose, D., & Myers, J. (2003). Measuring the Cost of State High School Exit Exams. In *Pay Now or Pay Later: The Hidden Costs of High School Exit Exams*. Center on Education Policy, Washington, DC. Retrieved from http://www.edlawcenter.org/ELCPublic/elcnews_090611_HiddenCosts.pdf.
- Rosenshine, B. (2003). *High Stakes Testing: Another Analysis*. University of Illinois at Urbana-Champaign, Urbana, IL. Retrieved from [http://faculty.ed.uiuc.edu/rosenshi/High stakes testing Another analysis.htm](http://faculty.ed.uiuc.edu/rosenshi/High%20stakes%20testing%20Another%20analysis.htm).
- Sadker, D., & Zittleman, K. (2004). Test Anxiety: Are Students Failing Tests Or Are Tests Failing Students? *Phi Delta Kappan*, 85(10), 740-751.
- Samuelson, S. (2001). Student Testing: The Stakes Are Rising. *State Legislatures*, 27(8), 30-34.
- Shepard, L.A. (2002). The Hazards of High-Stakes Testing. *Issues in Science and Technology*, 19(2). Retrieved from <http://www.issues.org/19.2/shepard.htm>.
- Stecher, B.M. (2002). Consequences of Large-Scale, High-Stakes Testing on School and Classroom Practice. In L.S. Hamilton, B.M. Stecher, & S.P. Klein (Eds.), *Making Sense of Test-Based Accountability in Education*. Santa Monica, CA: RAND.
- Steeves, K.A., Hodgson, J., & Peterson, P. (2002). Are We Measuring Student Success with High-Stakes Testing? *The Educational Forum*, 66(3), 228-235.
- Steinberg, L. (2003). Does High-Stakes Testing Hurt Students? *Education Week*, 22(21). Retrieved from <http://www.edweek.org/ew/articles/2003/02/05/21steinberg.h22.html>.
- Taylor, G., Shepard, L., Kinner, F., & Rosenthal, J. (2002). *A Survey of Teachers' Perspectives on High-Stakes Testing in Colorado: What Gets Taught, What Gets Lost*. Center for Research on Evaluation, Diversity and Excellence, University of California, Santa Cruz and Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles. ERIC Document Reproduction Service No. ED475139.
- Triplett, C.F., & Barksdale, M.A. (2005). Third Through Sixth Graders' Perceptions of High-Stakes Testing. *Journal of Literacy Research*, 37(2), 237-260.

- Triplett, C.F., Barksdale, M.A., & Leftwich, P. (2003). Children's Perceptions of High Stakes Testing. *Journal of Research in Education*, 13(1), 15-21.
- Viadero, D. (2003). Reports Find Fault with High-Stakes Testing. *Education Week*, 22(16). Retrieved from <http://www.edweek.org/ew/articles/2003/01/08/16stakes.h22.html?r=2034513868>.
- Vogler, K.E. (2002). The Impact of High-Stakes, State-Mandated Student Performance Assessment on Teachers' Instructional Practices. *Education*, 123(1), 39-55.
- Volante, L. (2004). Teaching To the Test: What Every Educator and Policy-Maker Should Know. *Canadian Journal of Educational Administration and Policy*, 35. ERIC Document Reproduction Service No. EJ848235.
- von der Embse, N.P. (2008). *High Stakes Testing, Socio-economic Status and Test Anxiety: Opportunities for Impact*. Thesis submitted to the Department of Educational Psychology, Miami University, Oxford, OH. Retrieved from http://etd.ohiolink.edu/send-pdf.cgi/von_der_Embse_Nathaniel.pdf?miami1214510534.
- Vu, P. (2008). Do State Tests Make the Grade? *Pew Research Center Publications*, January 17, 2008. Retrieved from <http://pewresearch.org/pubs/696/do-state-tests-make-the-grade>.
- Walker, S.F. (2000). High-Stakes Testing: Too Much? Too Soon? *State Education Leader*, 18(1), 1-3. ERIC Document Reproduction Service No. ED452219.
- Westchester Institute for Human Services Research. (2003). High-Stakes Testing. *The Balanced View*, 7(1). Retrieved from <http://www.sharingsuccess.org/code/bv/testing.pdf>.
- Wilde, M. (2008). *Are We Stressing Out Our Kids?* Retrieved from <http://www.greatschools.org/parenting/teaching-values/stressed-out-kids.gs?content=645>.
- Winters, M.A., Greene, J.P., & Trivitt, J.R. (2008). Building On The Basics: The Impact of High-Stakes Testing on Student Proficiency in Low-Stakes Subjects. *Civic Report No. 54*. Center for Civic Innovation at the Manhattan Institute, New York, NY. ERIC Document Reproduction Service No. ED502123.
- Wise, L.L., Sipes, D.E., Harris, C.D., Ford, J.P., Sun, S., Dunn, J., et al. (2002). *California High School Exit Examination (CAHSEE): Year 3 Evaluation Report*. Prepared by the Human Resources Research Organization for the California State Department of Education and California Department of Education, Sacramento, CA. Retrieved from <http://www.cde.ca.gov/ta/tg/hs/documents/y3exesum.pdf>.
- Wright, W.E. (2002). The Effects of High Stakes Testing in an Inner-City Elementary School: The Curriculum, the Teachers, and the English Language Learners. *Current Issues in Education*, 5(5). Retrieved from <http://cie.asu.edu/volume5/number5/>.
- Yeh, S.S. (2005). Limiting the Unintended Consequences of High-Stakes Testing. *Education Policy Analysis Archives*, 13(43). Retrieved from <http://epaa.asu.edu/epaa/v13n43>.
- Zellmer, M.B., Frontier, A., & Pheifer, D. (2006). What Are NCLB's Instructional Costs? *Educational Leadership*, 64(3), 43-46.

