

Teachers' Beliefs about the Effects of High Stakes Testing

Lantry L. Brockmeier¹, Robert B. Green¹, James L. Pate², Rudo Tsemunhu¹ & Michael J. Bochenko¹

Abstract

Since the enactment of the No Child Left Behind Act of 2001, high stakes testing has continued to be one of the major driving forces behind educational reform. In this study, Georgia teachers' beliefs about the effects of high stakes testing were examined. A random sample of teachers from 100 of Georgia's elementary schools, middle schools, and high schools responded to a 49-item survey measured on a five-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Items were grouped into six domains: curriculum, teaching, work satisfaction, stress, accountability, and students. Teachers' responses did not differ by gender, educational level, or school level. African American teachers responded more positively than White teachers on the survey. Teachers' positive and negative responses were discussed and recommendations were made for teachers and school leaders.

Keywords: accountability, high stakes testing, standardized testing, teachers

Kaback (2006) indicated that due to America's obsession with testing, high stakes standardized testing will not become an endangered species anytime soon. Although there has been some resistance in education circles during the last 10 years, the general public, policymakers, and parents continue to demand better school performance and view the results of high stakes testing as proof of learning (Scherer, 2005; Wahlberg, 2003). The results of high stakes tests may reveal to taxpayers that their investment is producing quality outcomes (Lederman & Burnstein, 2006).

As taxpayers, many parents want information allowing them to make comparisons of their children's and school's performance.

Driesler (2001) reported 90% of parents wanted information that would allow the comparison about their children and schools. Moreover, 83% of parents indicated that high stakes tests provide important information about their children's education. Poll and survey data have indicated a positive view of standardized testing by the general public (Phelps, 2005). Phelps indicated the percentage point differential between positive responses and negative responses to standardized testing varied from a +90% for students passing a graduation test, a +39% for ranking schools, and a +28% for promoting students to the next grade. In a recent survey of parents, Tompson, Benz, and Agiesta (2013) reported parents think standardized testing should be used to (a) ensure students meet adequate national standards (83%), (b) rank or rate schools (65%), (c) evaluate teacher quality (60%), and (d) determine whether or not students are promoted or can graduate (58%).

¹ Valdosta State University

² PhD, 1500 N. Patterson St., Valdosta, GA 31698. Email: jl pate@valdosta.edu, Phone: 229.333.5633

In addition, Tompson et al. indicated 75% of parents think standardized tests measure performance well or very well and 69% of parents indicated standardized tests measure the quality of schools well or very well.

Although standardized testing has been taking place in the United States since the 1840s (Resnick, 1982), the use of standardized testing really did not expand significantly until after World War II (Linn, Miller, & Gronlund, 2005). In an attempt to equalize educational opportunities the Elementary and Secondary School Act of 1965 was enacted. Along with the authorization, Congress required accountability for student progress and as a result testing increased. Concerns continued to grow over the quality of education in the 1970s (Airasian, 1988). From these concerns, minimum competency testing began in an attempt to ensure all students had mastered the basic skills (Hamilton, 2003).

Cizek (2001) noted the testing movement was likely due to poor decisions or perception of poor decisions by educators. Burton (1978) indicated the minimum competency testing movement transferred some important decisions from individual teachers to increased standardization of content, whereas Popham (1978) noted the minimum competency testing movement halted the devaluation of the high school diploma. Awareness between curriculum, instruction, and assessment grew as an attempt was made to ensure all students received at least the same minimum knowledge and skills in selected content areas (Camilli, Cizek, & Lugg, 2001).

From the minimum competency testing movement, the concept of measurement-driven instruction was fostered (Hamilton, 2003).

The prevailing thought was testing could influence what was taught. Concern over student and school performance continued to mount in the 1980s as *A Nation at Risk* was published (National Commission on Excellence in Education, 1983). Yet again, there was an increase in testing and school-level incentives (Hamilton, 2003). The standards movement in the 1990s led to increased links between standards, curriculum, and testing. Smith, O'Day, and Cohen (1990) indicated the links and formal stakes enhanced motivation to increase performance. High stakes testing encouraged educators and students to adopt a serious approach to teaching and learning (Lewis, 2000). The standards movement led to the No Child left Behind (NCLB) Act of 2001, which is a reshaped, reauthorization of the Elementary and Secondary School Act (Linn, Miller, & Gronlund, 2005). The NCLB Act consists of goals in the form of standards, tests or measures of performance, targets for performance, and consequences for a school's success or failure (Hamilton & Koretz, 2002).

The common denominator in educational reform over the past five decades has been the increased use of high stakes testing for accountability purposes due to the concern for student and school performance. Just as business leaders were concerned about students' ability to read and write during the 1970s (Cizek, 2001), researchers (Darling-Hammond, 2006; Haycock, 2005; Wakefield, 2003) today are lamenting that students are exiting school without the knowledge and skills to survive in an increasing competitive world.

Why high stakes testing? Phelps (2003) revealed grade point averages and course grades are too unreliable to be used as outcome measures, while Holland (2001) indicated standardized tests are essential to confirm grading systems varying from teacher-to-teacher and from school-to-school and to ensure standards are more than a suggestion. Other researchers (Fremer, 2005; Linn, Miller, & Gronlund, 2005) have indicated to argue against the use of test results dismisses relevant information leading to better decision making. Afflerbach (2005) offered three reasons why the general public commonly supports high stakes testing: fairness, scientific due to the tests undergoing examination for validity and reliability, and commonplace. A fourth reason for high stakes testing's popularity is the ability to provide a numerical score indexed to an alphabet representing quality and achievement (Baines & Stanley, 2004).

If one reads the educational literature on high stakes testing, one would get the impression of high stakes testing having few advantages since so much of the recent literature is negative (Stone, 2003).

However, teachers and schools routinely used standardized tests for documentation of student, teacher, and school performance during most of the 20th century. As long as the results or information was kept at the local level everything was fine. It was not until policymakers began to hold schools accountable for test results that the limitations became fatal flaws (Stone, 2003).

Teachers were very supportive of high stakes standardized testing in the 1970s and 1980s when the stakes were only for students (Phelps, 2005). In 1999, prior to the implementation of the NCLB Act, Phelps noted that there was a percentage point differential of +73% in teachers' support for standards, testing, and accountability. However, by the implementation of the NCLB Act of 2001, teachers' support for standards, testing, and accountability had declined to a percentage point differential of a +55%. While the response by teachers was still very positive, there had been a decline in support. Teachers are now under mounting pressure to increase performance of their students. The potential consequence of high stakes test results for teachers possibly include a negative evaluation, removal, reassignment, and a decrease in financial compensation (Herman, 2007; Hursh, 2005; Jones, 2001; Nichols & Berliner, 2007; Tashlik, 2010). The consequences of high stakes testing for teachers has increased further since the development of the value-added model in education and the potential use as part of a teacher's evaluation (Di Carlo, 2012; Linn, 2006; Milanowski, 2011; Newton, Darling-Hammond, Haertel, & Thomas, 2010; Paige, 2012; Ritter & Shuls, 2012; Scherrer, 2012; Wiliam, 2010). How are these potential consequences affecting teachers?

Purpose

The purpose of this study was to examine Georgia teachers' views about the impact of high stakes testing. Teachers were asked to respond to items crossing six domains: curriculum, teaching, work satisfaction, stress, accountability, and students. A secondary purpose was to determine if there were differences in teachers' responses about high stakes testing by (a) gender, (b) race or ethnicity, (c) educational level, and (d) school level.

Methodology

The methodology section is divided into two subsections. First, we will discuss the population, sample, and sampling procedure. Second, we will present information about the instrument.

Population, Sample, and Sampling Procedure

The Georgia Department of Education School Directory was used to select a stratified random sample of 100 elementary schools, middle schools, and high schools. Once schools were identified, teachers were randomly sampled from within school levels. The goal was to obtain sufficient data for a ± 5 percent margin of error at a 95 percent confidence level (i.e., 386 teachers). After two mailings, 300 teachers completed the survey. Due to incomplete data, 15 surveys eventually were removed from the data analysis. Although almost 78% of the desired responses were obtained, the actual response rate was approximately 38% due to oversampling in order to account for nonrespondents.

An examination was made to determine how closely the sample matched the statewide demographics. The number of teachers in the sample reporting to be at the elementary level and secondary level were similar to the statewide population, $\chi^2(1) = 2.74, p = .10$. The number of teachers in the sample reporting to be male and female were similar to the statewide population, $\chi^2(1) = 1.54, p = .22$. The number of teachers in the sample reporting to be Black, Hispanic, and White were not similar to the statewide population, $\chi^2(2) = 16.68, p < .001$. The sample had slightly fewer Black teachers and Hispanic teachers than the statewide population.

Demographic information collected on the survey included gender, race or ethnicity, educational level, and school level. The number and percentage of teachers responding to the survey by gender were 236 (83%) female teachers and 47 (17%) male teachers. By race or ethnicity, there were 239 (84%) Caucasian teachers, 37 (13%) African American teachers, 2 (0.75%) Hispanic teachers, 1 (0.35%) American Indian or Alaskan Native teacher, and 4 (1.41%) "Other" teachers. By educational level, 80 (28%) teachers reported having a bachelor's degree, 120 (42%) teachers reported having a master's degree, 67 (24%) teachers reported having an educational specialist's degree, and 16 (6%) teachers reported having a doctorate.

By school level, there were 154 (54%) teachers at elementary schools, 53 (19%) teachers at middle schools, 73 (26%) teachers at high schools, and 3 (1%) teachers at combination schools.

Instrumentation

The *Teacher's High Stakes Testing Survey* is a 49-item questionnaire constructed on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*) across six domains (curriculum, teaching, work satisfaction, stress, accountability, and students) with both positive and negative statements about high stakes testing. The instrument is a slightly modified version of the *Teacher's High Stakes Testing Survey* originally developed by Hope, Brockmeier, Lutfi, and Sermon (2006). Subsequently, Brockmeier, Pate, and Leech (2008) provided an in-depth analysis of the psychometric characteristics (i.e., validity and reliability) of the survey. An expert panel reviewed the instrument plus an exploratory factor analysis and confirmatory factor analyses were conducted. Expert panel members only suggested a few very minor modifications to improve the instrument. The confirmatory factor analyses yielded data to support the fit of the model and measurement invariance of the model by gender and race or ethnicity. In the present study, Cronbach's alpha reliability coefficient was .94 for the total scale. Reliability was very good for the total composite score. Negatively worded items were reverse-coded for the estimates of reliability and subsequent inferential statistical analyses.

Results

The results section consists of two subsections: analysis of items and inferential statistical analyses. First, an analysis of teachers' responses to items within subscales is reported using the median value and the percentage point differential (in parenthesis) between positive responses and negative responses for each item. Second, the results from the inferential statistical analyses are presented.

Analysis of Items

Beliefs about curriculum. The curriculum subscale consists of eight items. Teachers agreed (median value of 4) on 6 of 8 curriculum items (see Table 1).

Teachers agreed the (a) high stakes test content is aligned with the school's curriculum (+35%), (b) students' scores provide feedback to improve the curriculum (+31%), (c) high stakes testing requires teachers to teach to the test (+65%), (d) high stakes testing has led teachers to rethink about subject matter that is important to teach (+68%), (e) high stakes testing promotes some subject area content over other subject content (+83%), and (f) high stakes testing is counter to the idea of a balanced curriculum(+56%). Finally, teachers disagreed (median value of 2) that high stakes test items reflect the content students learn in a school's curriculum (-22%) and students' scores reflect the quality of a school's curriculum (-48%).

Table 1: Percentage of Responses and Descriptive Statistics by Item for Curriculum

Item		1 ^a	2	3	4	5	<i>Mdn</i>	<i>M</i>	<i>SD</i>
1	High stakes testing has led teachers to reassess their beliefs about subject matter that is important to teach.	5	7	8	52	28	4.00	3.92	1.03
2	High stakes testing is counter to the idea of a balanced curriculum (equal attention to subjects).	6	10	12	43	29	4.00	3.81	1.13
3	Students' scores on a high stakes test accurately portray the quality of a school's curriculum.	21	46	14	17	2	2.00	2.34	1.06
4	High stakes testing requires teachers to teach to the test.	2	10	11	39	38	4.00	4.00	1.05
5	High stakes test items accurately reflect the content students learn through a school's curriculum.	14	37	20	27	2	2.00	2.66	1.08
6	High stakes testing promotes certain subject area content over other subject area content.	2	3	7	47	41	4.00	4.20	0.88
7	Students' scores on a high stakes test provide feedback for schools to improve the curriculum.	5	20	19	49	7	4.00	3.33	1.03
8	High stakes test content is aligned with a school's curriculum.	4	17	24	46	10	4.00	3.41	0.99

Note. ^a 1 (Strongly Disagree), 2 (Disagree), 3 (Neither Agree nor Disagree), 4 (Agree), and 5 (Strongly Agree).

Beliefs about teaching. The teaching subscale consists of 11 items. Teachers agreed (median value of 4) with two items (see Table 2).

Teachers agreed that (a) high stakes testing requires test preparation that diminishes time to teach other subject content (+74%) and (b) high stakes testing reduces the teaching and learning process to a student's test score (+58%). Teachers neither agreed nor disagreed (median value of 3) that (a) high stakes testing motivates teachers to improve the teaching and learning process (-1%), (b) students' scores on a high stakes test provide information for teachers to improve their teaching (+16%), (c) high stakes testing has increased cooperation among teachers (-8%) and (d) high stakes testing has increased teacher and principal cooperation (-13%). Teachers disagreed (median value of 2) that (a) high stakes testing permits teachers to use the full range of their teaching skills (-49%), (b) high stakes testing leads to better teaching (-50%), (c) students' scores on a high stakes test are a valid measure of teaching ability (-72%), (d) the quality of a teacher's instruction is directly related to student performance (-62%), and (e) students' scores on a high stakes test are a valid way to determine the quality of education (-63%).

Beliefs about work satisfaction. Six items represent the work satisfaction subscale (see Table 3). Teachers agreed (median value of 4) with 5 of 6 items. Teachers agreed that (a) teacher satisfaction increases when she or he has input into the development of a high stakes test (+49%), (b) teacher work satisfaction decreases when the focus is on high stakes testing outcomes (+76%), (c) teachers leave low performing schools because of high stakes testing results (+48%), (d) high stakes testing diminishes the desire to be an educator (+58%), and (e) high stakes testing leads to teachers leaving the profession (+61%). Teachers strongly disagreed (median value of 1) that morale increased due to high stakes testing (-86%).

Table 2: Percentage of Responses and Descriptive Statistics by Item for Teaching

Item		1 ^a	2	3	4	5	<i>Mdn</i>	<i>M</i>	<i>SD</i>
9	High stakes testing permits teachers to use the full range of their teaching skills.	27	40	15	16	2	2.00	2.28	1.10
10	High stakes testing leads to better teaching.	25	40	20	12	3	2.00	2.28	1.05
11	Students' scores on a high stakes test are a valid measure of teaching ability.	41	39	12	7	1	2.00	1.87	0.93
12	Students' scores on a high stakes test are a valid way to determine the quality of education.	28	45	17	9	1	2.00	2.10	0.94
13	The quality of teachers' instruction is directly related to student performance on a high stakes test.	33	40	17	10	1	2.00	2.07	0.99
14	High stakes testing requires test preparation that diminishes time to teach other subject content.	3	6	8	49	34	4.00	4.05	0.96
15	Students' scores on a high stakes test provide information for teachers to improve their teaching.	9	22	22	43	4	3.00	3.11	1.09
16	High stakes testing reduces the teaching and learning process to a student's test score.	2	10	19	43	27	4.00	3.84	0.99
17	High stakes testing motivates teachers to improve the teaching and learning process.	15	21	28	32	3	3.00	2.88	1.11
18	High stakes testing has increased cooperation among teachers.	13	24	33	26	3	3.00	2.81	1.06
19	High stakes testing has increased teacher and principal cooperation.	13	27	33	23	4	3.00	2.78	1.06

Note. ^a 1 (Strongly Disagree), 2 (Disagree), 3 (Neither Agree nor Disagree), 4 (Agree), and 5 (Strongly Agree).

Table 3: Percentage of Responses and Descriptive Statistics by Item for Work Satisfaction

Item		1 ^a	2	3	4	5	<i>Mdn</i>	<i>M</i>	<i>SD</i>
20	Teacher morale has increased because of high stakes testing.	55	33	11	2	0	1.00	1.59	0.75
21	High stakes testing diminishes the desire to be an educator.	4	7	21	41	28	4.00	3.82	1.03
22	Teachers leave low performing schools because of high stakes test results.	2	7	34	37	20	4.00	3.66	0.96
23	The use of high stakes testing as a single measure to determine student achievement leads to teachers leaving the profession.	1	7	22	42	27	4.00	3.87	0.94
24	Teachers' work satisfaction diminishes when the focus is on high stakes testing outcomes.	2	5	11	45	38	4.00	4.13	0.90
25	Teacher satisfaction increases when she or he has input into the development of a high stakes test.	3	7	31	49	10	4.00	3.56	0.88

Note. ^a 1 (Strongly Disagree), 2 (Disagree), 3 (Neither Agree nor Disagree), 4 (Agree), and 5 (Strongly Agree).

Beliefs about stress. The stress subscale consists of 10 items. Teachers strongly agreed (median value of 5) with 7 of 10 items (see Table 4).

Teachers strongly agreed that (a) district supervisors' pressure to improve high stakes test scores increases teacher stress (+91%), (b) principals' pressure to improve high stakes test scores increases teacher stress (+91%), (c) teachers experience stress in the effort to maintain their school's accountability grade (+96%), (d) teachers' stress increases when the school's accountability grade declines (+92%), (e) teachers' stress increases when the school receives a failing grade (+95%), (f) punitive measures associated with high stakes testing increase teacher stress (+94%), and (g) teachers' stress increases with the public advertisement of a school's high stakes test results(+89%). Teachers agreed (median value of 4) that (a) teachers leave the profession from stress due to high stakes testing (+72%) and (b) high stakes testing leads to competition among teachers (+48%). Teachers neither agreed nor disagreed (median value of 3) upon the pressure to improve high stakes test scores may result in teachers cheating to improve scores (+3%).

Beliefs about accountability. The accountability subscale consists of seven items (see Table 5). Teachers agreed (median value of 4) that (a) high stakes testing has increased awareness of accountability issues (+79%), (b) teachers are more accountable because of high stakes testing (+41%), and (c) high stakes testing has increased teacher accountability for students' academic performance (+54%). Teachers disagreed (median value of 2) that (a) high stakes testing is a reform measure that improves the quality of education (-47%), (b) high stakes testing is an effective means for determining the quality of education (-64%), (c) high stakes testing promotes a cooperative environment between teachers and the community (-49%), and (d) students' scores on a high stakes test are an indicator of whether a school is staffed with high quality teachers (-64%).

Table 4: Percentage of Responses and Descriptive Statistics by Item for Stress

Item	1 ^a	2	3	4	5	<i>Mdn</i>	<i>M</i>	<i>SD</i>
26 High stakes testing leads to competition among teachers.	3	12	23	44	19	4.00	3.64	1.01
27 Teachers' stress increases when the school receives a failing grade.	1	1	2	32	65	5.00	4.60	0.63
28 Teachers' stress increases when the school's accountability grade declines.	1	1	4	37	57	5.00	4.49	0.68
29 Punitive measures associated with high stakes testing increase teacher stress.	0	1	4	35	60	5.00	4.53	0.64
30 Teachers experience stress in the effort to maintain their school's accountability grade.	1	0	2	36	60	5.00	4.55	0.63
31 Teachers' stress increases with public advertisement of a school's high stakes test results.	1	1	8	38	53	5.00	4.41	0.74
32 The pressure of high stakes testing may result in teachers cheating to improve scores.	12	24	25	27	12	3.00	3.03	1.22
33 District supervisors' pressure to improve high stakes test scores increases teacher stress.	0	1	6	37	55	5.00	4.46	0.70
34 Principals' pressure to improve high stakes test scores increases teacher stress.	1	1	4	38	55	5.00	4.45	0.74
35 Teachers leave the profession because of stress related to high stakes testing.	1	2	21	38	37	4.00	4.07	0.89

Note. ^a 1 (Strongly Disagree), 2 (Disagree), 3 (Neither Agree nor Disagree), 4 (Agree), and 5 (Strongly Agree).

Table 5: Percentage of Responses and Descriptive Statistics by Item for Accountability

Item	1 ^a	2	3	4	5	<i>Mdn</i>	<i>M</i>	<i>SD</i>
36 High stakes testing has increased teachers' accountability for students' academic performance.	4	10	17	52	16	4.00	3.67	1.01
37 High stakes testing has increased teachers' awareness of the accountability issues in education.	2	4	8	66	19	4.00	3.96	0.81
38 High stakes testing is an effective means of determining the quality of public education.	28	44	20	7	1	2.00	2.11	0.94
39 Students' scores on a high stakes test are an indicator of whether a school is staffed with high quality teachers.	33	40	18	8	1	2.00	2.05	0.96
40 High stakes testing is a reform measure that improves the quality of education.	24	35	29	11	1	2.00	2.30	0.99
41 Teachers are more accountable because of high stakes testing.	9	13	14	50	13	4.00	3.44	1.16
42 High stakes testing creates a cooperative environment between teachers and the community.	26	35	28	9	3	2.00	2.28	1.03

Note. ^a 1 (Strongly Disagree), 2 (Disagree), 3 (Neither Agree nor Disagree), 4 (Agree), and 5 (Strongly Agree).

Beliefs about students. The student subscale consists of seven items (see Table 6). Teachers strongly agreed (median value of 5) high stakes testing induces anxiety in students (+88%). Teachers agreed (median value of 4) that (a) high stakes testing contributes to students dropping out of school (+58%), (b) the pressure of high stakes testing may result in students cheating to improve scores (+63%), (c) teachers are concerned about the impact of high stakes testing on minority students (+75%), and (d) high stakes testing has changed the nature of student-teacher interactions (+58%). Teachers neither agreed nor disagreed (median value of 3) upon high stakes testing motivating students to achieve (-26%). Teachers strongly disagreed (median value of 1) that students' learning styles are accounted for in high stakes testing (-76%).

Table 6: Percentage of Responses and Descriptive Statistics by Item for Students

Item	1 ^a	2	3	4	5	<i>Mdn</i>	<i>M</i>	<i>SD</i>
43 High stakes testing contributes to the number of students that drop out of school.	2	7	24	45	22	4.00	3.79	0.93
44 Students' learning styles are accounted for in high stakes testing.	51	31	12	5	1	1.00	1.74	0.93
45 High stakes testing induces anxiety in students.	2	1	5	35	56	5.00	4.41	0.85
46 High stakes testing motivates students to achieve.	17	32	28	23	0	3.00	2.58	1.03
47 The pressure of high stakes testing may result in students cheating to improve scores.	2	7	19	49	23	4.00	3.85	0.93
48 Teachers are concerned about the impact of high stakes testing on minority students.	1	4	15	48	32	4.00	4.06	0.85
49 High stakes testing has changed the nature of student-teacher interactions.	2	9	20	43	26	4.00	3.82	0.98

Note. ^a 1 (Strongly Disagree), 2 (Disagree), 3 (Neither Agree nor Disagree), 4 (Agree), and 5 (Strongly Agree).

Inferential Statistical Analyses

To investigate questions about differences in teachers' responses by gender, race or ethnicity, educational level, and school level two independent means *t* tests and two ANOVAs were conducted on teachers' total scores ($M = 113.58$, $SD = 23.35$) on the instrument.

The teachers' total scores on the instrument were symmetric about the mean with a skewness value of -0.02 and with a kurtosis value of -0.20. Other statistical assumptions for the statistical procedures were met as well.

An independent means t test revealed there was no significant difference on the total score by teacher gender, $t(277) = 1.69, p = .09$. Since the race or ethnicity of the responding teachers' was primarily African American or White, an independent means t test was conducted. The independent means t test revealed there was a significant difference on the total score by race or ethnicity, $t(274) = 2.02, p = .04$.

Overall, African American teachers responded more positively ($M = 120.57, SD = 22.39, n = 37$) than White teachers ($M = 112.28, SD = 23.27, n = 239$) to items on the *Teacher's High Stakes Testing Survey*. Cohen's d , a measure of effect size, was 0.35 indicating a small to moderate practical effect. An analysis of variance revealed there was no significant difference on the total score by the teacher's educational level, $F(3,279) = 0.67, p = .57$. Finally, an analysis of variance revealed there was no significant difference on the total score by the school level of the teacher, $F(2,277) = 1.55, p = .21$.

Discussion

Teachers responded positively to a number of the survey items. Georgia teachers responding to the survey agreed that high stakes testing has increased their awareness of accountability issues and accountability for students' academic performance has increased. Teachers agreed that high stakes testing has led teachers to rethink about subject matter that is important to teach. In addition, teachers agreed the school's curriculum is aligned with the high stakes standardized test content and students' scores provide feedback in order to improve the curriculum. Moreover, teachers agreed that high stakes testing requires teachers to teach to the test. Teaching the curriculum is a positive, whereas teaching to specific items invalidates the inferences made from the test scores (Linn, Miller, & Gronlund, 2005; Popham, 2001, 2004). If teachers are teaching Georgia's statewide curriculum, then teachers are teaching the knowledge and skills deemed by the state as important and are teaching the material necessary to increase student academic performance.

One more positive finding was a teacher's work satisfaction increased when she or he participated in the development of Georgia's high stakes tests. These teachers work satisfaction might have increased due to having some feeling of control and a sense of responsibility, whereas teachers not participating in the tests' development may be frustrated lacking any type of control (Vogler & Virtue, 2007). In addition, teachers' assessment literacy may have increased due to participation in the development and as a result felt less stress and less anxiety about the test (Popham, 2009).

Overall, teachers neither agreed nor disagreed (+16%) that students' scores on a high stakes test provide information for teachers to improve their teaching. In fact, this might be true.

Linn (2001), Hamilton (2003), and Popham (1999, 2009) indicated most educational tests provide more global information about achievement of standards than the detailed information necessary for detailed diagnoses. While the timing of high stakes testing might not allow for immediate feedback to teachers for assisting individual students, within domains of knowledge teachers might determine strengths and weaknesses. Knowledge of overall student performance within a domain may provide sufficient feedback for a teacher to contemplate their content coverage of the standards as well as the teaching process. It was also interesting to note, despite teachers being held more accountable for student performance, teachers neither agreed nor disagreed (-1%) that high stakes testing motivates teachers to improve the teaching and learning process. In fact, responding teachers disagreed (-50%) that high stakes testing leads to better teaching. One possibility for this response is teachers' responding to the survey agreed (+74%) that high stakes testing requires test preparation which takes time away from teaching subject content.

Another possibility for this response is the pressure teachers feel from district supervisors (+91%) and from principals (+91%) to increase test scores.

While teachers and principals are being held more accountable for student achievement and school performance, one might think there would be an increased cooperation among teachers and increased teacher cooperation with their principal. However, this was not the perception of many teachers. Teachers neither agreed nor disagreed (-8%) that high stakes testing has increased cooperation among teachers. In fact, teachers actually agreed (+48%) that high stakes testing leads to competition among teachers. In addition, teachers neither agreed nor disagreed (-13%) with the statement that high stakes testing has increased teacher and principal cooperation. Teachers disagreed (-49%) that high stakes testing has created a cooperative environment between teachers and the community.

Teachers tended to disagree with a number of the survey items. Georgia teachers disagreed that high stakes testing is a reform measure designed to improve the quality of education or students' scores on a high stakes test are an indicator of whether a school is staffed with high quality teachers. Furthermore, teachers disagreed that students' scores on a high stakes test are a valid measure of teaching ability and the quality of a teacher's instruction is directly related to student performance. Most likely, teachers' responses to these items reflect their perception of control over the product (i.e., student performance on the high stakes test). It is doubtful teachers are abandoning responsibility, but the realization there are other factors impacting student performance of which teachers do not control. Prior student achievement and a host of other student, family, and community factors impact student performance on a high stakes test (Linn, 2006; Meyer, 2000).

Teachers disagreed (-49%) with high stakes testing permitting teachers to use the full range of their teaching skills. Additional investigation of responses to this question is warranted. Standards stipulate the goals, but do not mandate the curriculum or the instructional approach (Linn, Miller, & Gronlund, 2005). Researchers found no evidence of high stakes testing hindering the use of best teaching practices (Ayers, Sawyer, & Dinham, 2004), while others (Anderson 2009; Berube, 2004; Cook & Faulkner, 2006; Longo, 2010; Williamson, Bondy, Langley, & Mayne, 2005) called for using best teaching practices and teaching content knowledge over test preparation activities. Teachers in one study reported high stakes testing fosters creativity (Buck, Ritter, Jensen, & Rose, 2010). Moreover, Firestone, Monfils, and Schorr (2004) stated teachers with administrative support were more confident in using best teaching practices.

In general, teachers reported their work satisfaction decreased with high stakes testing. Teachers reported teachers are leaving low performing schools or the profession altogether due to high stakes testing and the pressure or stress to improve high stakes test scores. Teachers strongly agreed feeling self-inflicted pressure to improve high stakes test score results, but also felt pressure from the district supervisor and the school principal. Other researchers (Assaf, 2008; Pedulla, Abrams, Madaus, Russell, Ramos, & Miao, 2003; Taylor, Shepard, Kinner, & Rosenthal, 2003) have had similar responses to items about work satisfaction, stress, or pressure from high stakes testing. If educators and others responsible for public education had an adequate knowledge of assessment literacy (Popham, 2009), then it is very likely much of this perceived pressure would be reduced.

Teachers strongly agreed that high stakes testing induced anxiety in students. It is possible teachers' were transferring their anxiety onto students. In one study, 22% of parents reported testing was not at all stressful on their fifth grade children and only 24% parents reported testing was stressful for their children (Osburn, Stegman, Suitt, & Ritter, 2004). Fifth grade students, teachers, and counselors in another study responded to items about high stakes testing (Mulvenon, Stegman, & Ritter, 2005).

Teachers' thought testing created anxiety for students, but students generally reported no additional anxiety due to testing and some students even reported liking testing. However, high performing students occasionally reported anxiety over testing. Counselors reported teachers were much more stressed than students over testing.

Finally, teachers agreed that high stakes testing contributed to students dropping out of school. Hamilton (2003) aptly indicated it is difficult to determine if teachers' perceptions of more students dropping out is occurring. Some studies have reported an increased dropout rate (Darling-Hammond, 2003; McNeil, 2000; Pedulla et al., 2003) whereas other studies indicated there was no increase in the dropout rate or inconclusive evidence of increased dropout rates due to high stakes testing (Allensworth, 2005; Carnoy & Loeb, 2002; Warren & Jenkins, 2005).

Conclusion

Historically, one might make a case that high stakes standardized testing in the United States began around the mid-1840s when comparison of school performance and classroom performance started and soon afterwards became public knowledge (Linn, Miller, & Gronlund, 2005; Resnick, 1982). Certainly, the "high stakes" in high stakes standardized testing has shifted over time. Over the past five decades, there have been numerous educational initiatives increasingly relying on high stakes testing for accountability (Burton, 1978; Cizek, 2001; Hamilton, 2003; Linn, Miller, & Gronlund, 2005; Popham, 1978). Until recently, poll and survey questions responded to by the general public and educators were very positive in favor of high stakes testing. While poll and survey questions responded to by the general public are still very positive in favor of high stakes testing, poll and survey questions responded to by educators are less favorable since the passage of the No Child Left Behind Act (Phelps, 2005). It is not difficult to understand this decline in positive responses due to the increased pressure and consequences for teachers and administrators.

Overall, Georgia teachers responded similarly to items on the survey by school level, education level, and gender. However, African American teachers responded more positively than White teachers across survey items. While there was a statistical difference between these two groups, the effect size was a small to medium practical effect. The examination of teachers' responses to individual items revealed the tendency for Georgia teachers to respond negatively towards high stakes testing. This finding in itself was not surprising.

There are four recommendations for teachers. First, teachers should focus on teaching the statewide curriculum. Second, teachers should use best teaching practices for their content area rather than relying on test preparation activities in an effort to increase student performance on the statewide high stakes tests. Third, teachers should continue to learn more about assessment and become involved in the statewide test development process. As teachers increase their assessment literacy, teachers become more aware of the information that assessment results can and cannot provide to the practitioner. Finally, teachers should be careful not to transfer their own anxiety about statewide high stake testing to their students. If students become too anxious, then student performance may be affected negatively.

In closing, there are a number of recommendations for school leaders. First, school leaders should continue to increase their assessment literacy and the assessment literacy of teachers. School leaders should be knowledgeable about the strengths and limitations of statewide testing as well as the information that assessment results can and cannot provide to stakeholders. Second, school leaders should support teachers using best teaching practices over test preparation activities for increasing student performance. Third, while continuously striving to increase student performance in schools and school districts, school leaders should attempt to reduce the stress or pressure teachers feel as a result of high stakes testing in order to improve school culture. One way to reduce teacher stress and to improve school culture would be through the promotion of cooperation and collaboration among teachers and between school leaders and teachers. If successful, this in turn may positively impact student performance on the statewide high stakes tests.

References

- Afflerbach, P. (2005). National reading conference policy brief: High stakes testing and reading assessment. *Journal of Literacy Research*, 37, 151-163.
- Airasian, P. W. (1988). Symbolic validation: The case of state-mandated, high stakes testing. *Education Evaluation and Policy Analysis*, 10(4), 310-313.
- Allensworth, E. M. (2005). Dropout rates after high-stakes testing in elementary school: A study of the contradictory effects of Chicago's efforts to end social promotion. *Educational Evaluation and Policy Analysis*, 27(4), 341-364.
- Anderson, L. (2009). Upper elementary grades bear the brunt of accountability. *Phi Delta Kappan*, 90(6), 413-418.
- Assaf, L.C. (2008). Professional identity of a reading teacher: Responding to high stakes testing pressures. *Teachers and Teaching: Theory and Practice*, 14(3), 239-252.
- Ayers, P., Sawyer, W., & Dinham, S. (2004). Effective teaching in the context of grade 12 high-stakes external examination in New South Wales, Australia. *British Educational Research Journal*, 30 (1), 141-165.
- Berube, C.T. (2004). Are standards preventing good teaching? *The Clearing House*, 77, 264-267.
- Baines, L. A., & Stanley, G. K. (2004). High stakes hustle: Public schools and the new billion dollar accountability. *The Educational Forum*, 69(1), 8-16.
- Brockmeier, L. L. Pate, J. L., & Leech, D. W. (2008, November). Psychometric characteristics of the teacher's high stakes testing survey. Paper presented at the annual meeting of the Florida Educational Research Association, Orlando, Florida.
- Buck, S., Ritter, G. W., Jensen, N. C., & Rose, C. P. (2010). Teachers say the most interesting things- An alternative view of testing. *Phi Delta Kappan*, 91(6), 50-54.
- Burton, N. W. (1978). Societal standards. *Journal of Educational Measurement*, 15, 263-271.
- Camilli G., Cizek, G. J., & Lugg, C. A. (2001). Psychometric theory and the validation of performance standards: History, and future perspectives. In G. J. Cizek (Ed.), *Setting performance standards: concepts, methods, and perspectives* (pp. 445-475). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24, 305-331.
- Cizek, G. J. (2001). Conjectures on the rise and fall of standard setting: An introduction to context and practice. In G. J. Cizek (Ed.), *Setting performance standards: concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Cook, C. M., & Faulkner, S. A. (2006). Testing vs. teaching: the perceived impact of assessment demands on middle grades instructional practices. *RMILE Online*, 29, 1-13. Retrieved from http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/000001b/80/3e_/70/57.pdf.
- Darling-Hammond, L. (2003). Standards and assessments: Where we are and what we need. *Teachers College Record*. Retrieved from <http://www.tcrecord.org/Content.asp?ContentID= 11109>.
- Darling-Hammond, L. (2006). Constructing 21st-century teacher education. *Journal of Teacher Education*, 57, 300-314.
- Di Carlo, M. (2012). How to use value-added measures right. *Educational Leadership*, 38-42.
- Driesler, S. D. (2001). Whiplash about backlash. The truth about public support for testing. *NCME Newsletter*, 9(3), 2-5.
- Firestone, W. A., Monfils, L., & Schorr, R. Y. (2004). Test preparation in New Jersey: Inquiry-oriented and didactic responses. *Assessment in Education*, 11 (1), 67-88.
- Fremer, J. (2005). Foreword. In R. P. Phelps (Ed.), *Defending standardized testing*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Hamilton, L. S. (2003). Assessment as a Policy Tool. *Review of Research in Education*, 27, 25-68.
- Hamilton, L. S., & Koretz, D. M. (2002). Tests and their use in test-based accountability systems. In L. S. Hamilton, B. M. Stecher, & S. P. Klein (Eds.), *Making sense of test-based accountability in education* (pp. 13-49). Santa Monica, CA: RAND.
- Haycock, K. (2005). Choosing to matter. *Journal of Teacher Education*, 56, 256-265.

- Herman, J. L. (2007). *Accountability and assessment: Is public interest in K-12 education being served?* (CRESST Report No. 728). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Holland, R. (2001). *Indispensable tests: How a value-added approach to school testing could identify and bolster exceptional teaching*. Arlington, VA: Lexington Institute.
- Hope, W. C., Brockmeier, L. L., Lutfi, G. A., & Sermon, J. M. (2006, November). High stakes test's influence on teachers' beliefs. Paper presented at the annual meeting of the Florida Educational Research Association, Jacksonville, FL.
- Hursh, D. (2005). The growth of high-stakes testing in the USA: Accountability, markets and the decline in educational equality. *British Educational Research Journal*, 31(5), 605-622.
- Jones, L. V. (2001). Assessing achievement versus high-stakes testing: A crucial contrast. *Educational Assessment*, 7(1), 21-28.
- Kaback, S. (2006). High stakes are for tomatoes: Supporting teachers and students when the growing conditions are poor. *Kappa Delta Pi*, 42, 101-103.
- Lederman, L. M., & Burnstein, R. A. (2007). Alternative approaches to high stakes testing. *Phi Delta Kappan*, 87(6), 429-432.
- Lewis, A. (2000). *High-stakes testing: Trends and issues [Policy brief]*, Aurora, CO: Mid-Continent Research for Education and Learning.
- Linn, R. L. (2001). *The design and evaluation of educational assessment and accountability systems* (CSE Tech. Rep. 539). Los Angeles: Center for Research on Evaluation, Standards, and Student Testing.
- Linn, R. L. (2006). Validity of inferences from test-based educational accountability systems. *Journal of Personnel Evaluation Education*, 19, 5-15.
- Linn, R. L., Miller, M. D., & Gronlund, N. E. (2005). *Measurement and assessment in teaching* (9th ed.). Upper Saddle River, NJ: Pearson Education, Inc.
- Longo, C. (2010). Fostering creativity or teaching to the test? Implications of state testing on the delivery of science instruction. *The Clearing House*, 83, 54-57.
- McNeil, L. M. (2000). Creating new inequalities: Contradictions of reform. *Phi Delta Kappan*, 81, 729-734.
- Milanowski, A. (2011). Strategic measures of teacher performance. *Phi Delta Kappan*, 92(7), 19-25.
- Mulvenon, S. W., Stegman, C. E., & Ritter, G. (2005). Test anxiety: A multifaceted study on the perceptions of teachers, principals, counselors, students, and parents. *International Journal of Testing*, 5 (1), 37-61.
- Meyer, R.H. (2000). Value-added indicators: A powerful tool for evaluating science and mathematics programs and policies. *NISE Brief*, 3(3). Madison, WI; National Center for Improving Science Education, University of Wisconsin-Madison.
- National Commission on Excellence in Education. (1983). *A nation at risk*. Washington, DC: U.S. Department of Education.
- Newton, X. A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Education Policy Analysis Archives*, 18(23), 1-27.
- Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.
- Osburn, M., Stegman, C., Suitt, L., & Ritter, G. (2004). Parents' perceptions of standardized testing: Its relationship and effect on student achievement. *Journal of Educational Research & Policy Studies*, 4(1), 75-95.
- Paige, M. (2012). Using VAM. *Phi Delta Kappan*, 94(3), 29-32.
- Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Boston College: National Board on Educational Testing and Public Policy. Retrieved from <http://www.bc.edu/research/nbetpp/reports.html>.
- Phelps, R. P. (2003). *Kill the messenger: The war on standardized testing*. New Brunswick, NJ: Transaction Publishers.
- Phelps, R. P. (2005). Forty years of public opinion. In R. P. Phelps (Ed.), *Defending standardized testing*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

- Popham, W. J. (1978). As always, provocative. *Journal of Educational Measurement*, 15, 297–300.
- Popham, W. J. (1999). Why standardized tests? *Educational Leadership*, 8–15.
- Popham, W. J. (2001). Teaching to the test? *Educational Leadership*, 16–20.
- Popham, W. J. (2004). "Teaching to the test" An expression to eliminate. *Educational Leadership*, 82–83.
- Popham, W. J. (2009). Curriculum mistakes. *American School Board Journal*, 36–38.
- Resnick, D. (1982). History of educational testing. In A. K. Wigdor & W. R. Garner (Eds.) *Ability testing: Uses, consequences, and controversies* (pp.173-194). Washington, DC: National Academy Press.
- Ritter, G. W., & Shuls, J. V. (2012). If a tree falls in a forest, but no one hears ... *Phi Delta Kappan*, 94(3), 34–38.
- Scherer, M. (2005). Reclaiming testing. *Educational Leadership*, 63, 9.
- Scherrer, J. (2012). What's the value of VAM (value-added modeling)? *Phi Delta Kappan*, 93(8), 58–60.
- Smith, M. S., O'Day, J., & Cohen, D. K. (1990). National curriculum American style: Can it be done? What might it look like? *American Educator*, 14(4), 10–17, 40–47.
- Stone, J. E. (2003). Preface. In R. P. Phelps (Ed.), *Kill the messenger: The war on standardized testing*. New Brunswick, NJ: Transaction Publishers.
- Tashlik, P. (2010). Changing the national conversation on assessment. *Phi Delta Kappan*, 91(6), 55–59.
- Taylor, G., Shepard, L., Kinner, F., & Rosenthal, J. (2003). A survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught, what gets lost (CSE Tech. Rep. 588). Los Angeles: Center for Research on Evaluation, Standards, and Student Testing.
- Tompson, T., Benz, J., & Agiesta, J. (2013). Parents' attitudes on the quality of education in the United States. The Associated Press and NORC Center of Public Affairs Research.
http://www.apnorc.org/PDFs/Parent%20Attitudes/AP_NORC_Parents%20Attitudes%20on%20the%20Quality%20of%20Education%20in%20the%20US_FINAL_2.pdf
- Vogler, K. E., & Virtue, D. (2007). "Just the facts, ma'am": Teaching social studies in the era of standards and high-stakes testing. *Social Studies*, 98, 55–58.
- Wahlberg, H. J. (2003). Foreword. In R. P. Phelps (Ed.), *Kill the messenger: The war on standardized testing*. New Brunswick, NJ: Transaction Publishers.
- Wakefield, D. (2003). Screening teacher candidates: Problems with high-stakes testing. *The Educational Forum*, 67, 380–388.
- Warren, J. R., & Jenkins, K. N. (2005). High school exit examinations and high school dropout in Texas and Florida, 1971 - 2000. *Sociology of Education*, 78, 122–143.
- William, D. (2010). Standardized testing and school accountability. *Educational Psychologist*, 45(2), 107–122.
- Williamson, P., Bondy, E., Langley, L., & Mayne, D. (2005). Meeting the challenge of high-stakes testing while remaining child-centered. *Childhood Education*, 190–195.