

Running head: FAILING A HIGH STAKES TEST

Student Reactions to Being Wrongly Informed of Failing a High-Stakes Test:

The Case of the Minnesota Basic Standards Test

Dewey G. Cornell

University of Virginia

Jon A. Krosnick

LinChiat Chang

Stanford University

Under review, do not cite without permission

Author Notes. November 16, 2004. We thank the law firm of Larson, King and the Minnesota students and parents who participated in this study. Jon Krosnick is University Fellow at Resources for the Future. Address correspondence to Dewey G. Cornell, Ph.D. at the Programs in Clinical and School Psychology, Curry School of Education, University of Virginia, P.O. Box 400270, Charlottesville, Virginia 22904-4270. Email address dcornell@virginia.edu, or Jon A. Krosnick, Stanford University, 434 McClatchy Hall, Stanford, California 94305. Email address krosnick@stanford.edu.

Abstract

How do students react to being told that they have failed a test required for high school graduation? This study is based on an unfortunate event in 2000 when 7,989 students were wrongly informed that they had failed the Minnesota Basic Standards Test in Mathematics. We conducted a survey of 911 of these students to assess the psychosocial impact of this event. More than 80% of students reported that they increased studying, and nearly half made reductions in their extracurricular activities. Students reported a wide range of adverse emotional reactions, with over 80% reporting that they felt depressed, worried, or embarrassed. About half of the students said that they felt “stupid” and “less proud” of themselves. Fewer students reported troubling reactions by their parents or their peers. Adverse reactions were reported more frequently by girls than by boys and by students in higher grade levels. These findings point to the largely unaddressed need to study student reactions to failing high-stakes tests.

Student Reactions to Being Wrongly Informed of Failing a High-Stakes Test:

The Case of the Minnesota Basic Standards Test

High-stakes educational tests are increasingly used in the United States to make important educational decisions. Such tests often determine student placement into particular schools, course of study, promotion to the next grade, and graduation from high school (Heubert & Hauser, 1999). For example, eighteen states, which enroll half of all public school students in the country, currently require a high school exit exam or minimum competence test in order to receive a standard high school diploma (Center on Education Policy (CEP), 2002). During the next six years, at least 24 states will implement exit exams, affecting 70% of all public school students. In addition, the No Child Left Behind Act requires all states to test their students at least once between grades 10 and 12 (CEP, 2002).

Despite the tremendous national momentum toward high stakes testing, relatively little research has documented the socioemotional and behavioral impact on students when they are informed of test failure. What emotions are aroused by being told one failed a high stakes test? Does test failure motivate students to study harder and improve their test performance? How does it affect relationships with parents and peers? These questions are important, because many students are told they failed high stakes tests each year, and understanding their responses can guide efforts at remediation and perhaps prevent students from giving up and dropping out of school (Griffin & Heidorn, 1996; Huebert & Hauser, 1999).

An unfortunate accident in Minnesota schools provided researchers with the opportunity to study a group of students who had been informed of test failure. This study investigated the socioemotional and behavioral reactions of students who were *wrongly* informed that they had failed a high-stakes test in mathematics, the Minnesota Basic Standards Test. The survey was conducted at the request of a legal firm involved in a class action lawsuit on behalf of nearly

8,000 Minnesota students¹ who actually passed the test, but due to test scoring errors, received failing scores and hence were notified that they had failed. This study examines a disturbingly frequent event: mistaken notification of test failure (Rhoades & Madaus, 2003). The study also provides information that may be relevant to the broader group of students who actually fail high stakes tests.

Unfortunately, scoring errors are not rare; a national study by the National Board of Testing and Public Policy found 78 such errors, the majority (59) of which occurred since 1999 (Rhoades, & Madaus 2003). Test scoring errors affect millions of students; for example, a programming error by CTB McGraw Hill in converting reading comprehension raw scores to percentile scores on the TerraNova test affected a quarter of a million students in six states (Indiana, Nevada, New York, Tennessee, South Carolina, and Wisconsin). In New York City, 8,668 children were compelled to attend summer school and, in some cases, were not allowed to advance to the next grade, because the error gave them a score below the school division's cut-off. In Nevada, 736 students were told that they failed a test required for high school graduation when actually they had passed. Other large-scale test errors were detected in Arizona, California, Florida, Georgia, Massachusetts, Virginia, and other states.

This study contributes to our understanding of the impact of such events on students. The Minnesota students did not know that they had actually passed the test for several months, and during this time, they experienced the same consequences as students who actually failed; there is no reason for them to respond differently than students who had actually failed. And there was little difference in the mathematics skills of the students who were wrongly informed that they failed and other students who barely failed; the misinformed students who actually passed had true scores that were only slightly above the cut-off for passing. They failed the test because of scoring errors in only one or two questions.

We begin below by reviewing the current state of knowledge about the impact of failing a high-stakes test on students. Next, we describe the high stakes test we studied, the Minnesota Basic Standards Test, and the incident involving the test that gave rise to our survey. Finally, we describe the methodology employed to conduct the survey, report survey findings, and outline some of the implications for future educational and testing practice.

What Do We Know About The Impact Of Failing A High-Stakes Test?

In 1997, Congress directed the National Academy of Sciences (NAS) to study the use of testing for “student promotion, tracking, or graduation” (Public Law 105-78). The NAS report (Heubert & Hauser, 1999) examined trends in the use of high-stakes tests for all three purposes, whether such tests are used in a nondiscriminatory and legally defensible manner by school authorities, and specifically whether reading and mathematics tests yield valid information about student achievement. Notably, the NAS study did not examine the psychosocial ramifications of high-stakes testing on students or in particular the potential emotional impact of failing a high-stakes test.

Despite widespread and increasing use of high school exit exams, the NAS report concluded, “Very little is known about the specific consequences of passing or failing a high school graduation examination, as distinct from earning or not earning a high school diploma for other reasons....The consequences of using high-stakes tests to grant or withhold high school diplomas may be positive or negative. For example, if high-school graduation tests motivate students to work harder in school, the result may be increased learning for those who pass the test and, perhaps, even for those who fail” (p 288). Since a primary goal of high-stakes testing is to increase student achievement, it would seem important to study the effect of testing on student motivation. Why is so little known about the effects of testing on student motivation? Especially since students often have multiple opportunities to pass a high school graduation test, it would be

useful to know whether students (or which students) who fail an initial testing are in fact motivated to work harder in school so as to pass the test on the next occasion, or whether they are discouraged in a manner that decreases their effort and lowers subsequent achievement.

A longstanding and plausible assumption is that high-stakes testing increases student motivation to learn (Madaus, 1991; Madaus and Clarke, 2001), but even if this assumption is correct as a general principle, different forms of testing may have different motivational consequences, and there may be important individual differences in students' motivational response. Moreover, the motivational incentive to prepare for a high-stakes test can be distinguished from the motivational impact of test failure.

Much of the debate on the motivational impact of high-stakes testing has focused on racial and ethnic differences in test results (CEP, 2002). For example, Natriello and Pallas (2001) found that minority students were more likely than Caucasian students to fail graduation examinations in Texas, New York, and Minnesota. The Texas data further indicated that, among students who failed the examination on an initial attempt, African American and Hispanic students were less likely than white students to pass a subsequent examination and ultimately obtain a high school diploma. Such findings can be interpreted to suggest that test failure does not have equivalent motivational consequences across racial and ethnic groups (Natriello & Pallas, 2001). However, such inferences about student motivation are speculative; no investigation limited solely to data on group differences in test scores can demonstrate the influence of student motivation on test performance. Motivation is only one of many factors that might mediate the relation between initial test failure and subsequent performance in retaking a high-stakes test. Subsequent test performance can be influenced by the remedial resources and opportunities made available to students who failed an examination, and by the reactions of parents and peers who facilitate or inhibit remediation.

Another body of research has examined the impact of high stakes testing on drop-out rates. For example, Griffin and Heidorn (1996) investigated the relation between the Florida competency test administered first in the 10th grade (and up to five more times before high school graduation) and dropout behavior in 76,664 students in 75 high schools. Contrary to expectation, test failure did not contribute to an increased dropout rate among low achieving students or among minority students. However, test failure did appear to increase the dropout rate slightly in students with stronger academic records (higher GPAs), leading the authors to speculate that academically successful students might be more susceptible to “the perceived stigma attached to MCT failure.... [These] students may experience a substantial drop in self-esteem, or they may feel embarrassed in front of their peers. And such experiences might be especially acute for students with a proven record of academic success” (p 249, Griffin & Heidorn, 1996).

Many authorities in education and child development contend that high stakes testing can have an adverse emotional impact on students. Madaus (1991) compared the potentially hazardous effects of a proposed national test to the introduction of new drugs without careful determination of their safety. Among the disadvantages of high stakes testing, he listed “can be extremely stressful” and “can negatively affect such personality characteristics as self-concept and self-esteem” (p. 229, Madaus, 1991). Over a decade later, the Alliance for Childhood (2001) issued a position statement highly critical of high-stakes testing, endorsed by such well-known authorities as Robert Coles, David Elkind, Howard Gardner, Stanley Greenspan, Harold Howe II, Jonathon Kozol, Alvin Poussaint, Jerome Singer, TheodoreSizer, Thomas Sobol, and Robert Sternberg. Among other criticisms, the position statement asserted that high-stakes testing “may harm children’s health by causing excessive anxiety and stress.” Interviews with teachers and

parents in two states with high stakes testing seemed to concur with this opinion (Barksdale-Ladd & Thomas, 2000).

Despite the weight of expert opinion, the few studies that have directly investigated student reactions to graduation tests have generated inconsistent findings. One of the few such efforts is an unpublished study by Jonas and Hayes-Wallace (1986), who administered the Piers-Harris Self-Concept Test to tenth grade students before and after taking the Georgia Basic Skills Tests in mathematics and reading. Students who failed one or more of the tests showed a statistically significant decline in self-concept in one sample, but not in a replication sample the following year. Study findings were clouded by a high rate of participant attrition at posttest and by a test preparation program implemented for the replication sample that may have substantially reduced the number of students who failed.

Catterall (1989) surveyed an academically diverse sample of 736 ninth and eleventh grade students from eight states concerning their views of tests required for graduation. He found that 53% of students provided affirmative answers to the question, “Do tests required for graduation discourage some students from staying in high school?”, while 19% answered negatively and 28% were unsure. In contrast, only 14% stated that they were “personally aware of someone who failed a graduation test and later left school because of this.” Low achieving students were more likely than high achieving students to perceive that graduation tests were discouraging and to know someone who had left school as a result.

Nevertheless, Catterall found widespread student support for a graduation test requirement. Most students (56%) answered “yes” to the question, “Are tests required for graduation a good idea?”; only 20% responded “no,” and 24% were “not sure.” Among those who had passed a graduation test, support for the test was 72%; among those who initially failed and then later passed the test, support was 58%; and among those who had never taken such a

test, support was 50%. Even among those who had failed the test and had not yet passed it, support was 47%, with only 42% not supporting the test and 11% answering “not sure.”

Schmidt (1999) studied the stressful impact of repeatedly failing the Minnesota Basic Standards Tests on a purposeful sample of 16 students who would not have been expected to fail because of adequate academic records (GPA 2.0 or higher and California Achievement Test Total Battery above the 25th percentile). Based on interviews with students and parents, she concluded that test failure had damaged students’ confidence in their abilities, strained their relationships with parents and peers, and altered their educational and career plans.

Hughes and Bailey (2001) wrote about their experiences as high school teachers in Indiana, where students must pass tests in mathematics and language arts in order to graduate. The tests are first administered in the 10th grade and can be taken up to five times before graduation. The teachers reported that, to their surprise and dismay, students were not highly anxious about testing, and few 11th graders attended after-school preparation sessions, even though 38% of them had failed the exam the previous year. In discussions and interviews with students, the teachers found that students tended to dismiss the testing requirement as “unfair” and so concluded that it was “not worth worrying about.” Students observed that if they failed the test, they could take it again, and if they failed repeatedly, they could apply for a waiver or obtain a GED. This observation raises the possibility that attitudes may change as students move closer to graduation. As Madaus and Clarke (2001) pointed out, “there is no reason to believe that the motivational power of examinations will be the same at all grade levels” (p 97). Test anxiety might increase in the twelfth grade as the consequences of test failure become more imminent.

Perhaps the most novel study of high stakes testing was conducted by Wheelock, Bebell, and Haney (2000), who asked students to draw themselves taking the Massachusetts

Comprehensive Assessment System, a challenging series of tests that were administered over a period of days in grades 4, 8, and 10. The demanding nature of the tests triggered public outcry and protests from students, parents, and teachers. Not surprisingly, the most common themes portrayed in the drawings were negative responses, such as anxiety, anger, and pessimism, particularly among the older students. This study supports the familiar observation that students do not enjoy being tested, but it does not lend insight into the specific impacts of failure on those who fail.

Overall, there is much more rhetoric than direct evidence that graduation tests have a stressful impact on students (Barksdale-Ladd & Thomas, 2000; Langenfeld, Thurlow, & Scott, 1997; Wheelock, Bebell, & Haney, 2000). The available studies of student reactions are based on small samples and much of the information is anecdotal. The lack of research on the impact of failing a high school graduation test is puzzling. High stakes graduation testing is currently popular, but not new. A wave of states implemented minimum competency testing for high school graduation starting in the 1970's (Heubert & Hauser, 1999), and concerns have often been expressed about the stressful impact of such testing. Such concerns certainly merit investigation, since so many thousands of students fail such tests each time they are administered.

Indirect Evidence That High Stakes Testing May Be Stressful

Despite the paucity of direct evidence that failing high-stakes tests, and graduation tests in particular, have damaging effects on students, there is considerable evidence that test performance more generally is a source of anxiety and concern for students. Many studies have demonstrated the influence of test anxiety on student test performance from elementary school through post-secondary education, and much effort has been devoted to the development of test anxiety treatment (Zeidner, 1998; Spielberger & Vagg, 1995). We could find no study, however, of test anxiety as it relates to high school exit exams. Moreover, the study of test anxiety

concerns an anticipatory response; ironically, we know more about student *fears* of test failure and how to ameliorate them than student *reactions* to actual failure.

There is also a diverse literature on adolescent stress aimed at determining the influence of stress on various undesirable outcomes, such as substance use, teen pregnancy, family conflict, and emotional disturbance (Repetti, McGrath, & Ishikawa, 1999). Studies of adolescent stress routinely find that fear of academic failure is one of the most powerful and pervasive forms of stress that high school students experience (Armacost, 1989; deAnda, Baroni, Boskin, Buchwald, Morgan, Ow, Gold, & Weiss, 2000; Hurrelmann, Engel, Holler, & Nordlohne, 1988; McGuire, Mitic, & Neumann, 1987; Nordlohne & Hurrelmann, 1990; Repetti, McGrath, & Ishikawa, 1999). A study of a socioeconomically and ethnically diverse sample of 333 Los Angeles high school students (deAnda et al., 2000) found that among 76 different sources of stress, the stressors most commonly experienced by adolescents were “own expectations about career,” “future life plans,” “tests,” “grades,” “homework,” and “own expectations about school.” These academic stressors exceeded stressors concerning personal appearance, sexuality, conflicts with parents, alcohol and drug use, and peer relationships.

A Canadian study (McGuire, Mitic, & Neumann, 1987) of 1,684 students in grades 7-12 found that girls reported higher levels of stress than boys, but that both groups listed “schoolwork” as among the greatest sources of stress. Boys at each grade level consistently gave schoolwork the highest overall ranking, while for girls, schoolwork was second only to personal appearance as a perceived stressor. A West German study of 1,717 teenage students found that failure to meet parent expectations regarding academic achievement had a direct effect on health problems and subsequent use of medication (Hurrelmann, Engel, Holler, & Nordlohne, 1988; Nordlohne & Hurrelmann, 1990). The West German study is particularly relevant because the

West German educational system used high stakes testing to determine educational placement, which was regarded as critical to subsequent career options.

One more example is a study of suicidal behavior by de Wilde and Kienhorst (1998) that used a 77-item life stress scale to compare the experiences of 48 adolescents who attempted suicide with 66 depressed, non-suicidal adolescents and 43 non-depressed control adolescents. “Repeating a class” was one of three items (along with a change in living situation and a change in caretaker) occurring in the preceding year that distinguished adolescents who attempted suicide from both comparison groups.

The Minnesota Basic Standards Test.

In 1993, the Minnesota State Legislature directed the State Board of Education to establish a comprehensive set of requirements for high school graduation that included demonstrated proficiency in reading and mathematics. One outcome of this directive was that, beginning with the Class of 2000, all students must pass minimum competency tests in reading and mathematics, known as the Basic Standards Tests (BST). These two BSTs are regarded as covering material that should be mastered by 8th grade students. The tests are given first in the 8th grade and can be retaken in February and July of each subsequent year. For seniors who have not yet passed the test, there is a final opportunity to retake the tests in April. The mathematics and reading tests were first administered in February 1996. From 1998 to 2002, the initial pass rate for the mathematics test ranged from 70% to 74%, and for reading, the pass rate ranged from 68% to 80% (CEP, 2002).

In the summer of 2000, Minnesota educators learned that 47,097 students received incorrect scores on the February and April rounds of the mathematics BST because of errors in the scoring keys for two of the test forms (Minnesota Department of Children, Families, & Learning, 2000). As a result, 7,989 students in grades eight through twelve were notified that

they failed the test when actually they had passed.² The scoring problem was uncovered by a father whose daughter failed the mathematics test. In order to help his daughter prepare for the next examination, the father obtained a copy of the test and found errors in the scoring key. The testing company later admitted responsibility for the errors (Minnesota Department of Children, Families, & Learning, 2000).

In 2000, four class action lawsuits were filed against the testing company on behalf of the students who were wrongly told that they failed the test. In March 2002, a Minnesota district court granted the plaintiffs permission to conduct a joint survey of the affected students (Order Setting Trial Plan, Hennepin County District Court, March 25, 2002). The purpose of the survey was to gather systematic information on the impact of the test failure. The survey was carried out in May and June. In October 2002, just days before the trial was to begin, the parties reached a settlement that would provide up to seven million dollars in payments to the affected students. The settlement provided compensation to students based on the “adverse consequences” they experienced as a result of the scoring error, including the consequences of missing high school graduation ceremonies, attending summer school or remedial classes, and receiving mental health treatment for emotional distress. The court action recognized that being told that one failed the Minnesota Basic Standards Test could be a source of “significant emotional distress” that merited compensatory damages.

Assessment Of Impact On Students

The task of assessing the impact on students of being told that they failed the Minnesota Basic Standards Test was challenging and unprecedented. The two most significant challenges were that data could not be collected until approximately two years after the testing took place, and there were no established instruments to assess reactions to test failure. We also needed a

methodology that would collect data from a large sample of students in a relatively short period of time.

We therefore chose to employ the most commonly-used method of survey data collection today: telephone interviewing. Telephone interviewing has a number of procedural advantages, including the ability to locate geographically diverse individuals quickly and efficiently, and to present tailored sequences of questions in which answers to a question determine which questions are asked subsequently (using computer-assisted telephone interviewing software). Unlike paper and pencil surveys, telephone interviews require no literacy skills for respondents (because all questions and answers are expressed orally), and interviewers can be sure that respondents answer all questions (rather than accidentally skipping a question on a paper-and-pencil questionnaire) (see Weisberg, Krosnick, & Bowen, 1996). For a study of this sort, it would have been much more time-consuming and expensive to send interviewers out to conduct face-to-face interviews with the respondents. Mailing out self-administered paper-and-pencil questionnaires would likely have entailed a lengthier field period to achieve the same response rate. Therefore, telephone interviewing seemed to be the most practical approach.

We relied on retrospective accounts of the experience of failing the test. The existing literature on memory and recall of distant events identifies potential measurement problems with such data (e.g., Stone, Turkkan, Bachrach, Jobe, Kurtzman, & Cain, 2000). Self-reports may be susceptible to biases in recall and self-presentation, although these problems are thought to be more prominent in more distant childhood recollections or in recall of traumatic events. Because the students we interviewed had learned that they did pass the test prior to the interviews, some of the stigma of failure that might have influenced their willingness to talk about a personal shortcoming was no longer present. Finally, we were cognizant of a body of research on the diagnosis of psychiatric disorders that indicates that interviewees can reliably recall semantic

memories of generalized feelings and reactions to stress, even if they cannot recall precise episodes and factual details such as dates (e.g., Kessler, Wittchen, Abelson, & Zhao, 2000).

Focus groups and supplementary interviews. To help develop survey questions appropriate to student knowledge and recall of the test failure experience, we observed three focus groups of students who discussed their reactions to being told that they failed the test. The focus groups were conducted at the facilities of a firm called Delve in Minneapolis. Delve staff recruited the participants in the focus groups by telephoning them and offering them a fee to participate in a group discussion on the topic of standardized testing in Minnesota.

The individuals who participated in the focus groups were selected from lists of 100 students who had been high school seniors in 2000 and another 100 students who had been in grades 8 to 11 in 2000. These 200 students were randomly selected from the complete lists of affected students who lived within a limited geographic range around Delve's office, to permit practical transportation to the focus group meetings. Two of the focus groups were composed of students who had been in grades 8 to 11 in 2000. The third focus group was composed only of students who had been seniors in 2000, based on the possibility that these students might have been the most strongly affected by their test experience, since they were prevented from graduating with their classmates.

The focus groups were moderated by Professor Robert Mitchell of Clark University, and the researchers observed the discussions from behind a one-way mirror. Dr. Mitchell facilitated a general conversation about the Minnesota Basic Standards Test and then asked a series of open-ended questions concerning participants' experiences with the BST in 2000 and how they responded to the initial incorrect report that they had failed.

The focus group participants talked openly about their reactions to being told they failed the test and offered many detailed recollections about their experiences. Being told they failed

the test appeared to be a salient event for all the students, even though there were obvious differences in how strongly the students reacted to the test failure information. Some students sharply curtailed their extracurricular activities in order to spend additional time studying for the next exam, while other students were relatively unconcerned and made few changes in their daily lives. For some students, the event appeared to be mildly disappointing, while for others, particularly seniors who had hoped to graduate with their classmates, test failure was an event of traumatic proportions. These observations indicated that students could recall their experiences and confidently describe how they felt and reacted to the news that they failed the test.

Students were not as able to recall details such as the length of time during which they felt sad or depressed about failing the test, or the number of occasions in which they were teased by siblings or peers. However, students were confident that they could recall whether or not they had such feelings or experiences. Students could also recall significant behavioral events, such as whether they quit a part-time job or an extracurricular activity in order to increase time studying for the exam, and whether they attended summer school or were placed in a remedial mathematics class. We calibrated the level of detail required in subsequent survey questions accordingly.

We also adapted our survey to match the terms and phrases that students used to describe their experiences. Ordinarily, we would have been exceedingly reluctant to use a survey question that bluntly asked students if failing the test made them “feel stupid,” but in fact the term “stupid” was widely used by the students in the focus groups and therefore seemed suitable for use in the survey.

After the focus groups, we identified five topics to address in the survey: academic impact, extra-curricular activities, emotional reactions, social relations with peers, and social relations with family. First, we wanted to know the academic impact, which included what the

students did to prepare to take the exam again. We knew that many of the students took summer school or remedial classes because school divisions offered these opportunities specifically for students who had failed the test. We also learned that many students spent extra time studying on their own or with the help of parents, teachers, or friends. A common complaint among students was that in order to take summer school or to do extra studying for the exam, they had to reduce or give up other activities, such as part time jobs and participation in sports, so the second area of inquiry was impact on extracurricular activities.

The third domain was students' emotional response to being told they failed the test. Students in the focus groups variously described themselves as depressed, discouraged, embarrassed, or in some other way troubled by the news that they had failed. Closely related to the comments about their emotional response were participants' perceptions of how others responded to them. Accordingly, the fourth area we asked about was the responses of peers, because some students attempted to keep secret from their peers that they failed the test, and others reported being teased or looked down upon by their peers. The fifth area involved the responses of parents, since we anticipated that parent reactions would be of great concern to students who failed a test that was required for high school graduation.

After the survey was completed, we had several opportunities to interview students who participated in the class action suit and their parents. Respondents included some students who had been identified by attorneys as suffering substantial adverse effects of testing, so that they were probably not representative of all the affected students. To balance this potentially skewed group, we contacted some additional students at random and invited them and their parents to participate in interviews as well. Some of these students, especially the younger students who would have had multiple opportunities to retake the examination, described much less distress

associated with the news of test failure. These interviews generated a wealth of anecdotal accounts that supplemented the survey data.

Study Questions And Hypotheses

We set out initially simply to describe the types of experiences that students had in response to the news that they had failed the test. Expecting to find variation across students in their experiences, we posited two hypotheses that might explain some of this variation. First, we thought that test failure might have had greater impact on students in higher grade levels, because the consequence of not graduating from high school would be more imminent for these individuals, and the number of opportunities to retake the examination would have declined. For high school seniors, failing the examination might be especially stressful. Second, we hypothesized that there might be gender differences in students' emotional reactions: boys may react less emotionally and more stoically than girls to such traumatic experiences (Barrett, Lance, Sechrest, & Schwartz, 2000; Dindia & Allen, 1992).

Method

Data Collection

The telephone survey was conducted by Schulman, Ronca, and Bucuvalas, Inc. (SRBI), a highly respected survey research firm based in New York City. SRBI employs closely supervised interviewers who gain extensive experience working on a full range of telephone survey projects. Interviewing for this project began in early May, 2002, and continued into June, 2002.

Each telephone number in the sample was called up to 20 times on different days and at different times of the day in order to attempt to contact a person. When a call was answered, the interviewer asked to speak with the student by name; if the student was not available, the interviewer asked when would be a good time to call to reach the student. If the student could no

longer be reached at the telephone number, the interviewer asked for a different telephone number at which the student could be reached.

Students were offered \$10 to participate in the interview, which lasted approximately 10 minutes. If the student was a minor, the interviewer asked a parent for permission to speak to the student before initiating that conversation.

Students were told that SRBI was “conducting brief interviews with Minnesota students as part of an important survey about education in Minnesota” under the direction of a university professor. As is standard survey practice, the interviewers were not informed that the survey was being conducted as a part of a lawsuit or who was funding the survey.

Participants

The survey sample of 911 students was drawn from an initial sampling frame consisting of 7,973 students who were informed erroneously that they had failed the Mathematics Test of the Minnesota Basic Standards Test in February or April, 2000, minus 659 students for whom contact information was not provided by their school districts, four seniors who participated in a focus group, and 237 students who said they did not want to be included in the affected class for the lawsuit. Thus, the revised sampling frame included 7,077 students.

Seniors constituted a special impact group because they would have been prevented from graduating from high school and presumably would have experienced greater stress and disruption than other students. Because seniors constituted a small proportion of all affected students, we designated all 480 seniors with contact information as eligible to participate in the survey. A sample of 1,908 students was drawn randomly from the complete list of all 8th through 11th graders in the sampling frame. As a result, the total number of students identified for potential telephone contact was 2,388.

The American Association for Public Opinion Research (AAPOR) published a document in 1998 entitled *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for RDD Telephone Surveys and In-Person Household Surveys*. AAPOR (1998) defines the “cooperation rate” of a survey as “the proportion of cases interviewed of all eligible units ever contacted.” This is an indication of the extent to which contacted potential respondents refused to be interviewed. The cooperation rate for our survey was 85.7% among students who were reached by an interviewer. If the households that were contacted but refused to allow the interviewer to speak with the student are included in the cooperation rate calculation, the cooperation rate was 80.7%. Both of these are excellent cooperation rates for contemporary telephone surveys (Visser, Krosnick, & Lavrakas, 2000; Weisberg, Krosnick, & Bowen, 1996).

AAPOR (1998) defines “response rate” as “the number of complete interviews with reporting units divided by the number of eligible reporting units in the sample (p. 3)” and provides a series of mathematical formulas for calculating response rates in different ways. A total of 980 interviews were completed for our survey. The total number of student names released to the survey interviewers for possible interviewing was 2,388. Therefore, the most stringent response rate that could be calculated for this survey is $980 / 2388 = 41.0\%$. The bulk of the non-responses were not attributable to refusals, as indicated by the high cooperation rates for the survey. Instead, the bulk of non-responses occurred because listed students were not reached in order to attempt to begin an interview.

Many of the students listed in the State of Minnesota file did not have listed telephone numbers, or their listed phone numbers were not accurate. Extensive efforts were made to obtain correct phone numbers for as many students as possible in order to generate representative survey results. First, a list of all sampled students with address information (4,789 students) was sent to a company called *Tele-match* for directory look-up in order to find new or replacement

phone numbers. *Tele-match* returned 2,243 matched cases and 853 new phone numbers. 3,350 names were also sent to a firm called *National Change of Address* (NCOA) to find new addresses and phone numbers. NCOA returned 401 matching addresses and 110 new phone numbers. Then, the names associated with 970 dead phone numbers were sent to *Tele-match* for more intense manual look-up. This search generated 68 new phone numbers. 337 names of seniors were sent to attorneys involved in the case for look-up, and they obtained 91 new phone numbers. Finally, letters were mailed to 281 seniors who had not yet been reached; the letter offered an incentive of \$25 and provided recipients with a toll free telephone number to call to be interviewed. As a result of this mailing, 33 seniors called in to complete interviews. Sixty-six letters were returned as “undeliverable/bad address.”

Despite these extensive search procedures, we were unable to locate a valid telephone number for 749 students in the selected sample. If the response rate for the survey is recalculated after eliminating these individuals from the denominator because they *could not* be interviewed, the result is $980 / 1639 = 59.8\%$. In other words, interviews were completed with approximately 60% of the individuals for whom a telephone number could be located after attempting to locate phone numbers (or respondents) via the Court-ordered contact information list plus 5 different supplementary methods and a letter mail-out.

Of the possibly valid telephone numbers, 174 did not yield any contact with a potential respondent *after 20 attempts per respondent* because of technological impediments (e.g., call blocking, answering machines), no answer after 20 instances of calling the number, a respondent who was deceased or otherwise unreachable, or a respondent with health problems, inability to hear, or inability to speak and understand English sufficiently. If the response rate for the survey is again recalculated eliminating these people from the denominator because they *could not* be interviewed, the result is $980 / 1465 = 66.9\%$.

As can be seen in Table 1, the obtained sample of students closely matched the population of students in terms of year of birth, grade level, gender, and zip code. The sample also closely matched the population in their original, erroneous, test scores, and in the change in test scores when their test was properly scored.

Questionnaire Construction

The questionnaire was programmed for Computer-Assisted Telephone Interviewing (CATI); questions were displayed one at a time on computer screens, and when interviewers recorded answers by pushing keys, the program automatically displayed the next appropriate question.

The questionnaire was constructed to meet the highest standards of survey research methodology and to yield unbiased questions that would not tilt responses in favor of one side or another (see Visser, Krosnick, & Lavrakas, 2000). Survey question topics were selected because of their potential relevance to the legal questions under consideration, so some information that might be of interest from a research perspective (e.g., student racial or ethnic background) was not collected.

To design the questionnaire, we began with a list of the experiences that had been mentioned by the focus group participants and wrote questions to measure the frequency with which people had had each of those experiences as the result of being told incorrectly that they did not pass the math BST in 2000. Each question was carefully worded to avoid artifacts that might compromise the accuracy of people's reports of their behaviors and experiences. For example, all questions employed fully balanced wording, so that equal emphasis was placed on answer choices suggesting that individuals had undesirable experiences and on answer choices suggesting they did not have such experiences (see Schuman & Presser, 1981). Because the survey was conducted on behalf of the plaintiffs and could be subject to charges of bias to

promote the reporting of adverse effects of the test experience, answer choices were presented so that if any slight order effects were present, they would bias respondents away from reporting an adverse effect.

The words employed throughout the questionnaire were relatively simple and are used relatively often in everyday conversation. We avoided jargon whenever possible. Each question asking about the impact of the misinformation on students began with a phrase such as “As a result of being told incorrectly that you didn’t pass the test, ...”. The purpose of this phrase was to be sure that respondents clearly understood the referent of each question and did not misinterpret a question as referring to conditions other than those caused by the misinformation. For example, if a question simply asked “Did you work with a teacher after school during the regular school year?”, an affirmative answer might refer to work done for reasons other than having been told they failed the BST.

Questionnaire Content

The first section of the questionnaire set the stage for questions about the impact of test failure. Initial questions verified that the student had taken the Minnesota BST in math in February or April of 2000, had been told that he or she failed the test, and was later informed that he or she actually had passed the test.

Next, the interviewer asked a series of questions about the academic impact of failing the test, starting with whether the student took a summer school course to prepare to retake the mathematics BST:

“The next questions I’d like to ask are about what may or may not have happened to you as a result of being told you failed the test when in fact you had passed. Please think back to the weeks and months after you were told that you failed the test during the spring of

2000. As a result of being told incorrectly that you didn't pass the test then, did you go to summer school, or did you not go to summer school for that reason?"

The interview continued with questions about the time and cost involved in attending summer school. There were questions about other activities to prepare to retake the test, such as taking a remedial class during the regular school year and studying with a parent, teacher, or tutor after school hours. Students were asked if their family incurred any costs, such as paying for a tutor or enrolling in summer school. Students also were asked if they reduced their hours at a paying job or spent less time with friends as a result of being told they failed the test. Students who were seniors at the time they took the test were asked questions about their high school graduation plans and whether they attended their high school graduation ceremony.

Another series of questions asked students about the emotional impact of being told that they failed the test. Students were asked if they felt less interested in going to school as a result of being told they failed the test. They were asked if they felt depressed right after being told they did not pass the test, and if the answer was affirmative, to indicate how depressed they felt on a four-point scale (slightly, moderately, very, or extremely depressed). Students then answered a parallel set of questions about feeling "worried" and "embarrassed" in reaction to the news that they failed the test and rated the degree they had this feeling on the same four-point scale. Next, students were asked (yes or no) if being told they did not pass the test made them "feel stupid." The final two questions in this domain asked students if the news made them wonder if they were not as good at math as they thought they were, or to feel less proud of themselves.

The next series of questions asked students about their perceptions of their parents' reactions to the test result. Four questions asked students if their parents were disappointed in

them, were mad at them, respected them less, and took away privileges because they did not pass the test.

Three questions assessed peers: the interviewer asked respondents if other students found out that they did not pass the test and whether these students made fun of them or respected them less for not passing the test. The interviewer then asked if the respondents had brothers or sisters, and if so, whether their siblings teased them for not passing the test.

Results

Table 2 displays the raw frequencies of boys' and girls' responses to the 23 questions, grouped into five topical domains. To describe how many students were affected in some way and how many students endorsed multiple items, we present cumulative frequencies for each domain in Table 3. For example, only 13.5% of students reported no academic impact of being told they failed the test; 20.1% indicated one form of impact; 23.7% indicated two forms of impact; and 42.7% indicated three or more forms. The overwhelming majority of students reported some form of impact in the academic, extracurricular, and emotional domains. 86.5% of students reported some academic impact, and 48.8% reported some extracurricular impact.

Emotional impact was divided into two subsets because three of the items required a yes/no response, while three other items elucidated responses along a five-point rating scale. For the yes/no items (felt stupid, felt not as good as other students, felt less proud of self), 75% indicated at least one of these forms of impact. For the five-point scale items (felt depressed, worried, or embarrassed), 81.8% indicated at least one of the three forms of impact, and nearly one-third (32.8%) endorsed at least one form as greater than moderately present.

Few students reported parent or peer impact. 43.1% indicated one or more of the four forms of parent impact, and 24.3% indicated one or more of the three forms of peer impact.

These findings for peers should be considered in light of the results from a survey question that

asked, “Did any student at your school find out that you did not pass the test, or did no students at your school find out?” More than one-third (36.9%) of students reported that no student at their school knew that they failed the test.

In order to aggregate items into domain scores, we standardized each item, then summed items within each domain, and finally, standardized the domain scores. This procedure permits comparisons across domains by gender and grade level that are not affected by differences in the number of items or the number of response categories for different items. We did not expect that items within domains would be highly intercorrelated, since some responses might represent alternative reactions. For example, a student who was placed in a remedial math class might be less likely to attend an after-school test preparation program or take summer school as a result. However, for our purposes, endorsement of any of these actions represented a measurable impact of test failure.

We tested for gender and grade level effects by conducting a multivariate analysis of covariance (MANCOVA) on the five standardized domain scores with gender as an independent variable and grade level as a covariate. Means and standard deviations of standard scores are presented in Table 4, and multivariate and univariate results are presented in Table 5. For gender, there was a statistically significant F approximation to Wilks’s Lambda, $F(5, 904) = 12.595, p < .000, \eta^2 = .065$. For grade, the F approximation to Wilks’s Lambda also was statistically significant, $F(5, 904), p < .000, \eta^2 = .029$. Univariate F tests revealed statistically significant gender effects for both academic and emotional impact, but not for extracurricular, parent, or peer impact. There was a statistically significant effect of grade for academic, extracurricular, and emotional impact, but not parent or peer impact. The impact of the experience was consistently greater for older students, and girls reported more academic, extracurricular, and emotional impact than did boys.

Discussion

Many students were distressed to be told that they failed a high stakes test. Although the extent of impact varied across students, the cumulative findings indicate that the majority of students in our sample were substantially affected by purported test failure. Students increased their studying and made changes in their extracurricular activities in order to take the test again. Students reported multiple negative emotional reactions, such as feeling depressed or feeling stupid, and in some cases endured unpleasant responses by their parents and their peers. To our knowledge, these are the first published findings regarding the socioemotional impact of being told one failed an examination required for high school graduation.

Academic and extracurricular impact. In the academic domain, most students reported additional efforts to study in preparation for another opportunity to take the exam, which might be considered a positive and constructive response. Nearly two-thirds of students reported extra studying on their own, and substantial numbers (30-50%) worked with teachers, family members, or other adults -- in many cases pursuing more than one of these strategies. Nearly one-third of students took a summer school class, and one-quarter took a remedial class during the school year. As might be expected, the impact increased steadily from 8th to 12th grade.

The increased studying by students often meant a reduction in extracurricular activity. A small proportion of students (15.7%) reduced hours at a paid job, and a sizable number (39.3%) spent less time with friends. About one-quarter of the students (23.7%) changed summer vacation plans in order to attend summer school or undertake additional studying. We did not survey students about decreased involvement in school-sponsored clubs and sports, but recommend doing so in future studies. After the surveys were conducted, we had opportunities to learn anecdotally about the experiences of individual students who participated in the class

action lawsuit. Some students dropped out of school organizations, or quit football or basketball teams, in order to focus their efforts on passing the examination.

4.4% of students dropped out of school as a result of failing the examination. It is difficult to compare this figure to official Minnesota drop-out rates, because state calculations are based on a different age group of students and a definition of drop-out that was not employed in the survey we conducted. Nevertheless, since the students in our study did not actually fail the examination, their decision to drop out of school is especially troubling and deserves further investigation. Anecdotally, we know of several students who dropped out of school because they were discouraged and humiliated by failing the test. By the time they learned that they had actually passed the test, they had taken jobs and changed residences and ultimately decided not to return to school. These students also reported lingering feelings of resentment toward school authorities that affected their motivation to return to school. Clearly, news of test failure can have a strong impact on some students and might be an important influence on school drop-out rates.

Emotional impact. We anticipated that it would be difficult to gauge the full emotional impact of test failure on students, because they were surveyed almost two years after being told that they failed test and because they had since learned that they actually passed the test. Although the intensity of their feelings might well have dissipated, our focus groups and anecdotal observations persuaded us that most of these students had vivid recollections of their emotional reactions to being informed that they failed the test. The survey results indicated that more than three-fourths of the students reported adverse emotional reactions. For example, half of the students were willing to admit that they “felt stupid.” More than half reported that they “felt depressed” and more than half reported that they felt “embarrassed.” Most students reported multiple negative reactions.

As expected, girls reported greater emotional impact than boys. We do not know if the impact was truly greater on girls, because research on gender differences yields extensive evidence that females generally disclose more about their emotions than males (Dindia, & Allen, 1992). There is also evidence that females may have greater awareness of their emotional states and a capacity to describe their experiences in more detail (Barrett, Lane, Sechrest, & Schwartz, 2000). Nevertheless, both boys and girls reported adverse reactions to test failure, and both genders demonstrated increasing impact with grade level.

Anecdotally, we learned of students who felt shocked and humiliated to the point of feeling physically ill, with symptoms such as nausea and vomiting. Some students felt sufficiently depressed and discouraged to seek professional counseling. Other students reported intense feelings of anxiety and panic as they contemplated taking the exam again. The literature on test anxiety (Spielberger & Vagg, 1995) makes it clear that such symptoms are not rare, but there appears to be far more research on anticipatory test anxiety than the after-effects of test failure. The increasing prevalence of high stakes testing in the United States signals a need to expand the scope of test anxiety research.

Parental and peer impact. We observed less parental impact than expected. Relatively few students reported that their parents were mad at them (15.7%), respected them less (7%), or took away privileges (14.9%) in reaction to the news that they had failed the test. Just more than a third (37.7%) of students reported that their parents were disappointed in them. We wondered if these findings reflected a non-punitive and supportive approach to parenting among these families. We know that many parents constructively assisted their students in preparing to retake the exam by helping them to study or by providing them with additional study opportunities, such as summer school and professional tutoring. Another interpretation is that the parents had low regard for the testing process and assigned greater responsibility to the school system than to

their children for test failure. We spoke with parents who contended that the schools did not adequately prepare students for the examination or that high stakes testing was not an appropriate educational requirement. It would be useful to examine the role of parental attitudes on student preparation and success in high stakes testing in a prospective study.

Students also reported little negative reaction from peers. Few students reported being made fun of (11.6%) or respected less (8.5%) by peers because they failed the test. Only about one in eight students (14.7%) recalled being teased by a sibling. Anecdotally, we encountered two explanations from students about peer reactions. Among the students in lower grades, where the test failure rate is high, some students told us that they did not feel stigmatized by failure because many of their friends also failed. However, by the 11th and 12th grades, most students have passed the examination, so that failure is more likely to be a source of embarrassment. Students at this level who anticipated negative peer reactions may have decided not to disclose their test failure to peers. We are aware of several students who went to considerable lengths to pretend that they passed the examination. For example, one student held a long-planned graduation party and did not disclose to anyone that he was not eligible to graduate.

Study Limitations

Ideally, a study of student reactions to high stakes testing would assess a large random sample of students using a well-researched, standardized set of measures. Measures of study habits and involvement in extracurricular activities, academic self-concept and attitudes toward school, emotional adjustment, and family and peer relations would be collected prospectively before and after testing. We might include measures obtained from parents as well as students. We would examine changes in students over time and compare students who failed the test with students who passed the test. We would want to control for differences in academic aptitude, perhaps by contrasting students who failed by a small margin with students who passed by a

small margin. We would want information on student race, ethnicity, and social class, which was not available within the legal constraints of this study.

Of course, it was not possible to conduct a controlled, prospective study employing a pre-post design with this sample of students. Our task was to gather information on a group of students nearly two years after they had been informed that they failed the Minnesota BST. Under these circumstances, we undertook a survey that was carefully constructed and administered under standardized conditions to obtain student recollections of their reactions to their testing experience. The survey was limited to issues about which students seemed able to recall after such a long time interval. Extraordinary effort was made to locate as many students as possible and to assure that the sample was representative of those who had been misinformed that they failed the math portion of the BST.

One important feature of this sample is that the students did not actually fail the BST, all of them passed by a small margin. These circumstances distinguish them from students who genuinely failed the examination, but there is no reason to think their reactions to test failure would be any less substantial. There is an advantage in surveying students who later learned that they did not fail the test, since the stigma of failure was removed, so they might have been willing speak about the experience with less inhibition. In any case, the fortuitous opportunity to interview these students may lead the way for greater attention to the broader group of students who fail high stakes examinations.

Unfortunately, there may be other opportunities to study students who have been misinformed that they failed a high stakes exam. The regrettable experience of 7,989 Minnesota students is not an isolated incident. According to the report of the National Board on Educational Testing and Public Policy (Rhoades & Madaus, 2003), scoring errors on standardized tests are a pervasive problem which has come to public attention primarily because of the adverse

consequences when such tests are used to make high stakes decisions. The pervasiveness of scoring errors is cause for concern and reflects an iatrogenic effect of the high stakes testing movement that deserves careful consideration. In response to the testing error that wrongly prevented students from attending their high school graduation ceremonies in Minnesota, the State Education Commissioner wrote, “I can’t imagine a more horrible mistake that NCS could have made. And I can’t fathom anything that NCS could have done that would have caused more harm to students. I know that NCS agrees with me” (cited in Rhoades & Madaus, 2003, p 13).

The Need For Immediate Study Of High Stakes Test Failure

Although the phrase “more research is needed” is a familiar cliché in the closing sections of journal articles, the need for more study seems particularly acute in this case. High stakes testing affects millions of students, and current educational trends and politics make it likely that eventually it will affect every student in the United States.

The American Psychological Association (2001) has called for more research on the impact of high-stakes testing. “Because the stakes are so high for so many students, additional research should begin immediately to learn more about the intended and unintended consequences of testing in educational decision making. If tests are going to be used to determine which students will advance and what subjects schools will teach, it is imperative that we understand ... how the use of high-stakes testing will affect student drop-out rates, graduation rates, course content, levels of student anxiety, and teaching practices.”

The AERA position statement on high-stakes testing (AERA, 2000) does not specifically address the question of the stressful impact of testing or test failure, but does require “full disclosure of likely negative consequences of high-stakes testing programs.” AERA also states that, “Where credible scientific evidence suggests that a given type of testing program is likely to have negative side effects, test developers and users should make a serious effort to explain these

possible effects to policy makers.” The statement calls upon governmental authorities to conduct “ongoing evaluation of both intended and unintended consequences of high-stakes testing.”

Because students are offered multiple opportunities to pass minimum competency tests, it is important to examine the impact of initial test failure on students and the process by which they prepare for subsequent examinations. Research that considers only ultimate outcomes such as graduation or dropping out is not going to provide insight into ways to help students achieve ultimate success in the testing process.

Why has there been so little research on the socioemotional consequences of test failure? One possible reason that authorities have not investigated the stressful impact of test failure is that test failure is a ubiquitous and largely unavoidable consequence of any rigorous test. All students cannot reasonably be protected from test failure any more than they can be protected from other disappointments in life, such as losing an athletic contest or not being admitted to a favored college. The consequences of test failure became more salient in Minnesota, however, because the tests were scored in error, and students were subjected to an undeserved experience of failure. An analogy might be made to the pain that follows surgery. If the surgery is medically necessary and the stakes are high enough, the pain is an acceptable consequence, but if the surgery is undertaken because of a faulty diagnosis, then the pain is grounds for a grievance.

We do not know enough about the pain that follows the procedure of high-stake testing. We do not know which students are most vulnerable to what kinds of pain and how that pain might be lessened. And as any surgeon will tell you, some degree of post-surgical pain is good, because it prevents the patient from engaging in injurious activity before the damaged area has healed. How much pain is tolerable and healthy?

Test failure is and arguably should be a distressing experience for any student who aspires to succeed. Apathy would be a more troubling response than distress. Our findings are

not presented as a condemnation of high-stakes testing, only an argument that the painful consequences of test failure may be substantial and deserve consideration. We know too little about student reactions to test failure and the most effective means by which to help students cope with their distress and engage in adaptive responses that increase their likelihood of passing the test the next time.

There are compelling practical reasons to study student reactions to test failure that should be endorsed and supported by the proponents of high stakes testing. Students almost always have multiple opportunities to pass minimum competency tests, so it is important that they respond to failure in an adaptive manner and undertake preparations that are likely to result in success the next time they take the test. If students are discouraged by the failure, develop negative self-perceptions, or engage in self-defeating behavior by failing to study, their chances of success diminish.

We believe more attention should be given to the process of notifying and counseling students about their test results. We encountered a wide range of student reactions and experiences. Many students reported that they were so ashamed of failing the test that they attempted to keep it secret from their classmates. Some of these students were embarrassed later to be placed in remediation classes or to hear their name called by loudspeaker to attend a meeting for students who had failed the test. One student went so far as to pretend that he passed the test and graduated; two years later he has not been able to come to terms with the deception of his friends and relatives. Another student intercepted the mail to her parents and prevented them from learning that she had failed the test.

On the other hand, some parents and teachers were equally sensitive to test failure. Several parents we interviewed purposely avoided or delayed telling their son or daughter the test result had come in the mail because they anticipated that their children would be too

distressed by the news. One student tried unsuccessfully to find out from her teachers whether she had passed the test and later learned that they avoided speaking with her because they could not bear to break the bad news to her.

Perhaps there should be guidelines for counseling students who have failed a high stakes test and their parents. The mere reporting of a pass/fail outcome is not sufficient when the stakes are truly high and the student's response is so important to future academic success. Because minimum competency testing is so similar in content and form to standardized achievement testing, there seems to have been little differentiation of the test result reporting process.

Testing authorities have a clear responsibility to assure the accuracy of test scoring so that students are not subjected to the stress of test failure if they did not in fact fail the test. More broadly, educators have a responsibility to recognize the potentially adverse emotional impact of test failure on students and to take steps that will help students deal with high stakes test failure in a constructive and adaptive manner. Both the American Psychological Association (2001) and the American Educational Research Association (2000) have called for study of the effects of high stakes testing on students. We hope this study helps to stimulate a response to this call.

References

- Alliance for Childhood (2001). *High stakes testing position statement*. Retrieved October 4, 2001 from www.allianceforchildhood.net/news/histake_test_position_statement.htm.
- American Educational Research Association. (2000). AERA Position Statement Concerning High-Stakes Testing in PreK-12 Education. Retrieved January 19, 2003 from www.aera.net/about/policy/stakes.htm.
- American Psychological Association (2001). Appropriate use of high-stakes testing in our nation's schools. Retrieved January 19, 2003 from www.apa.org/.pubinfo/testing.html.
- 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, 384-397.
- Barrett, L.F., Lane, R., Sechrest, L., & Schwartz, G. (2000). Sex differences in emotional awareness. *Personality and Social Psychology Bulletin*, 26, 1027-1035.
- Catterall, J. (1989). Standards and school dropouts: A national study of tests required for high school graduation. *American Journal of Education*, 98, 1-34.
- Center on Education Policy. (2002). *State high school exit exams: A baseline report*. Washington, D.C.: Author. Retrieved January 10, 2003 from www.cep-dc.org.
- de Anda, D., Baroni, S., Boskin, L., Buchwald, L., Morgan, J., Ow, J., Gold, J.S., & Weiss, R. (2000). Stress, stressors and coping among high school students. *Children and Youth Services Review*, 22, 441-463.
- de Wilde, E.J., & Kienhorst, I.C.W.M. (1998). Life events and adolescent suicidal behavior. In Miller, Thomas W. (Ed.) *Children of trauma: Stressful life events and their effects on children and adolescents* (pp. 161-178). Madison, CT: International Universities Press.
- Dindia, K., & Allen, M. (1992). Sex differences in self-disclosure: A meta-analysis. *Psychological Bulletin*, 112, 106-124.

- Heubert, J.P., & Hauser, R.M. (Eds.) (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.
- Hughes, S., & Bailey, J. (2001). What students think about high-stakes testing. *Educational Leadership, 59*, 74-76.
- Hurrelmann, K., Engel, U., Holler, B., & Nordlohne, E. (1988). Failure in school, family conflicts, and psychosomatic disorders in adolescence. *Journal of Adolescence, 11*, 237-249.
- Jonas, E.D., & Hayes-Wallace, L. (1986) Minimum competency testing: An analysis of student outcomes for those not mastering mandated testing standards. (ERIC Document Reproduction Service No. ED246092)
- Langenfeld, K., Thurlow, M., & Scott, D. (1997). *High stakes testing for students: Unanswered questions and implications for students with disabilities*. (Synthesis Report No. 26). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Madaus, G. (1991). The effects of important tests on students: Implications for a National Examination System. *Phi Delta Kappan* (November), 226-231.
- Madaus, G., & Clarke, M. (2001). The adverse impact of high-stakes testing on minority students: Evidence from one hundred years of test data. In G. Orfield and M. L. Kornhaber (eds.). *Raising standards or raising barriers? Inequality and high-stakes testing in public education* (pp. 85-106). New York: Century Foundation Press.
- McGuire, D.P., Mitic, W, & Neumann, B. (1987). Perceived stress in adolescents: What normal teenagers worry about. *Canada's Mental Health, 35*, 2-5.
- Minnesota Department of Children, Families, & Learning (2000). Answers to anticipated district questions regarding scoring errors on February and April Mathematics Basic Standards

- Tests. Unpublished report. Minneapolis, MN: Author. Retrieved January 13, 2003 from <http://cfl.state.mn.us/testindex.html>.
- Natriello, G., & Pallas, A.M. (2001). The development and impact of high-stakes testing. In G. Orfield and M. L. Kornhaber (eds.). *Raising standards or raising barriers? Inequality and high-stakes testing in public education* (pp. 19-38). New York: Century Foundation Press.
- Nordlohne, E., & Hurrelmann, K. (1990). Health impairment, failure in school, and the use and abuse of drugs. *Health hazards in adolescence*, 8, 149-166.
- Schuman, Howard, and Stanley Presser. 1981. *Questions and Answers In Attitude Surveys: Experiments on Question Form, Wording, and Context*. New York: Academic Press.
- Repetti, R.L., McGrath, E.P., Ishikawa, S.S. (1999). Daily stress and coping in childhood and adolescence. In A.J. Goreczny & M. Hersen (Eds.) *Handbook of pediatric and adolescent health psychology* (pp 343-360). Needham Hts, MA: Allyn & Bacon.
- Rhoades, K., & Madaus, G. (2003). *Errors in standardized tests: A systematic problem*. Boston: National Board of Educational Testing and Public Policy. Retrieved September 3, 2003 from www.bc.edu/nbetpp.
- Spielberger, C.D., & Vagg, P.R. (1995). *Test anxiety: Theory, assessment, and treatment*. Washington, DC: Taylor & Francis.
- Stone, A.A., Turkkan, J.S., Bachrach, C.A., Jobe, J.B., Kurtzman, H.S., & Cain, V.S. (2000). *The science of self-report: Implications for research and practice*. Mahwah, NJ: Erlbaum.
- Visser, P. S., Krosnick, J. A., & Lavrakas, P. (2000). Survey research methods. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social psychology*. New York: Cambridge University Press.

Weisberg, H., Krosnick, J. A., & Bowen, B. (1996). *Introduction to survey research, polling, and data analysis*. Thousand Oaks, CA: Sage.

Wheelock, A., Bebell, D.J., & Haney, W. (2000). What can student drawings tell us about high-stakes testing in Massachusetts? *Teachers College Record*. ID Number: 10634. Retrieved January 25, 2002 from <http://www.tcrecord.org>.

Zeidner, M. (1998). *Test anxiety: The state of the art*. NY: Plenum.

Footnotes

¹Cornell and Krosnick were retained as expert witnesses by one of the law firms representing the plaintiffs.

²According to a report by the Minnesota Department of Children, Families, and Learning, 112,208 students took either Form A or Form B of the Mathematics BST in February, 2000. Six questions were scored incorrectly on Form B, affecting 45,739 students in grades 8 through 12. A total of 7,930 of these students were told they failed the test when actually they had passed. A total of 518 seniors were told incorrectly that they had failed the test, but many of them retook the test during a special administration for seniors in April. A total of 1,358 seniors took the April, 2000, version of the Mathematics BST, but this test had one question formatted incorrectly. After this question was dropped from consideration, fifty-nine students who had originally been told they failed the test achieved passing scores. After the April testing, 336 seniors remained who were incorrectly classified as failing the test, even though they had passed it on one or both occasions.

Table 1. Comparison Of The Survey Sample To The Population Of Students Who Were Told They Failed The Minnesota Basic Standards Test In Mathematics

	Percentages			Discrepancy	
	Full Population	Sampling Population	Survey Completions	From Full Population	From Sampling Population
Year of Birth					
1982 or earlier	17.5%	17.3%	20.2%	2.7%	2.9%
1983	20.8%	20.9%	17.1%	-3.7%	-3.8%
1984	33.5%	33.6%	31.8%	-1.7%	-1.8%
1985 or later	28.2%	28.3%	30.9%	2.7%	2.6%
N	7,701	7,296	910	Average: 2.7%	2.8%
Grade Level					
8 th grade	5.0%	5.0%	5.4%	0.4%	0.4%
9 th grade	44.9%	45.0%	46.7%	1.8%	1.7%
10 th grade	26.6%	26.7%	23.8%	-2.8%	-2.9%
11 th grade	16.4%	16.3%	10.1%	-6.3%	-6.2%
12 th grade	7.1%	6.9%	14.0%	6.9%	7.1%
N	7,737	7,330	911	Average: 3.6%	3.7%
Gender					
Female	50.1%	50.1%	52.7%	2.6%	2.6%
Male	49.9%	49.9%	47.3%	-2.6%	-2.6%
N	7,735	7,328	911	Average: 2.6%	2.6%
Original Test Score					
570-580	12.6%	12.5%	17.2%	4.6%	4.7%
583-584	14.7%	14.7%	14.3%	-0.4%	-0.4%
585-586	19.8%	19.7%	19.3%	-0.5%	-0.4%
587-589	0.7%	0.7%	1.2%	0.5%	0.5%
590	18.3%	18.5%	17.3%	-1.0%	-1.2%
592	17.8%	17.9%	18.9%	1.1%	1.0%
595	16.1%	16.0%	11.9%	-4.2%	-4.1%
N	7,737	7,330	911	Average: 1.8%	1.8%
Change in Test Score when Correctly Scored					
15 points or fewer	19.8%	19.8%	19.8%	0.0%	0.0%
16-17 points	21.7%	21.8%	21.1%	-0.6%	-0.7%
18-19 points	22.7%	22.8%	24.3%	1.6%	1.5%
20-22 points	35.7%	35.6%	34.9%	-0.8%	-0.7%
N	7,737	7,330	911	Average: 0.8%	0.7%
Zip Code Clusters					
50308 – 55100	-	12.1%	11.8%	-	-0.3%
55101 – 55300	-	14.1%	12.5%	-	-1.6%
55301 – 55400	-	14.3%	16.6%	-	2.3%
55401 – 55500	-	18.6%	16.2%	-	-2.4%
55501 – 56000	-	13.5%	12.9%	-	-0.6%
56001 – 56300	-	12.1%	13.4%	-	1.3%
56301 - 58103	-	15.3%	16.6%	-	1.3%
N	-	6,694	881	Average: -	1.4%

Table 2. Survey Response Frequencies For Boys And Girls Who Were Told They Failed The Minnesota Basic Standards Test In Mathematics

Item content	Boys N=431	Girls N=480	All Students N=911
Academic Impact			
1. Took summer school	26.5%	32.1%	29.4%
2. Worked with a teacher	30.2%	28.8%	29.4%
3. Worked with a family member	41.5%	52.1%	47.1%
4. Worked with a non-family member	29.9%	35.0%	32.6%
5. Extra studying on own	62.4%	67.9%	65.3%
6. Took a required class	24.8%	23.8%	24.3%
7. Dropped out of school	6.3%	2.7%	4.4%
Extracurricular Impact			
8. Reduced hours at paid job	17.2%	14.4%	15.7%
9. Changed summer plans	19.3%	27.7%	23.7%
10. Spend less time having fun with friends	36.0%	42.3%	39.3%
Emotional Impact			
11. Felt depressed			
Not at all	52.2%	36.7%	44.0%
Slightly	25.1%	25.8%	25.5%
Moderately	13.9%	21.9%	18.1%
Very	6.0%	9.6%	7.9%
Extremely	2.6%	5.0%	3.8%
12. Felt worried			
Not at all	36.0%	20.2%	27.7%
Slightly	22.3%	21.0%	21.6%
Moderately	26.0%	27.9%	27.0%
Very	11.4%	21.7%	16.8%
Extremely	4.4%	8.5%	6.6%
13. Embarrassed			
Not at all	54.3%	37.5%	45.4%
Slightly	17.6%	17.1%	17.3%
Moderately	14.6%	20.8%	17.9%
Very	9.0%	16.5%	13.0%
Extremely	4.4%	7.9%	6.3%
14. Felt stupid	40.8%	58.5%	50.2%
15. Wonder if not as good at math	57.3%	65.6%	61.7%
16. Felt less proud of self	41.8%	57.3%	49.9%
Parent Impact			
17. Parents disappointed in you	38.3%	37.1%	37.7%
18. Parents mad at you	15.5%	15.8%	15.7%
19. Parents respected you less	8.4%	5.8%	7.0%
20. Parents took away privileges	14.8%	15.0%	14.9%
Peer Impact			
21. Other students made fun of you	14.8%	8.8%	11.6%
22. Other students respected you less	9.0%	7.9%	8.5%
23. Teased by brother or sister	12.8%	16.5%	14.7%

Table 3. *Cumulative Frequencies For Boys And Girls Who Were Told They Failed The Minnesota Basic Standards Test In Mathematics*

Cumulative frequency for items	Boys N=431	Girls N=480	All Students N=911
Academic Impact			
Students reporting no academic impact	14.6%	12.5%	13.5%
Students reporting one item	20.6%	19.6%	20.1%
Students reporting two items	27.1%	20.6%	23.7%
Students reporting three or more items	37.6%	47.3%	42.7%
Students reporting at least one item	85.4%	87.5%	86.5%
Extracurricular Impact			
Students reporting no extracurricular impact	54.1%	48.5%	51.2%
Students reporting one item	26.2%	26.3%	26.2%
Students reporting two items	13.0%	17.5%	15.4%
Students reporting all three items	6.7%	7.7%	7.2%
Total students reporting at least one item	45.9%	51.5%	48.8%
Emotional Impact (depressed, worried, embarrassed)			
Students reporting no emotional impact	24.8%	12.3%	18.2%
Students reporting no impact greater than slight	47.6%	30.2%	38.4%
Students reporting no impact greater than moderate	76.6%	57.1%	66.3%
Students reporting one item greater than moderate	12.3%	22.3%	17.6%
Students reporting two items greater than moderate	7.2%	11.3%	9.3%
Students reporting three items greater than moderate	3.7%	7.9%	5.9%
Total students reporting at least one item	75.2%	87.7%	81.8%
Total reporting at least one item greater than moderate	23.2%	41.5%	32.8%
Emotional Impact (stupid, not as good, less proud)			
Students reporting no impact	29.9%	20.6%	25.0%
Students reporting one item	25.1%	16.3%	20.4%
Students reporting two items	20.2%	24.2%	22.3%
Students reporting all three items	24.8%	39.0%	32.3%
Total students reporting at least one item	70.1%	79.4%	75.0%
Parent Impact			
Students reporting no impact	55.9%	57.7%	56.9%
Students reporting one item	23.2%	21.7%	22.4%
Students reporting two items	11.8%	12.9%	12.4%
Students reporting three items	6.0%	4.6%	5.3%
Students reporting all four items	3.0%	3.1%	3.1%
Total students reporting at least one item	44.1%	42.3%	43.1%
Peer Impact			
Students reporting no impact	75.9%	75.6%	75.7%
Students reporting one item	13.9%	17.7%	15.9%
Students reporting two items	7.9%	4.6%	6.1%
Students reporting all three items	2.3%	2.1%	2.2%
Total students reporting at least one item	24.1%	24.4%	24.3%

Table 4. *Comparisons Of Boys And Girls Across Grades 8-12 In Five Impact Domains, Using Standardized Domain Scores*

	Boys		Girls		All Students	
	Mean	N	Mean	N	Mean	N
Academic Impact						
8 th graders	-.38	37	-.30	12	-.34	49
9 th graders	-.13	197	-.03	227	-.08	424
10 th graders	-.07	95	-.04	121	-.05	216
11 th graders	-.05	45	.39	47	.17	92
12 th graders	.20	57	.46	73	.33	130
Extracurricular Impact						
8 th graders	-.06	37	.05	12	-.01	49
9 th graders	-.07	197	.00	227	-.04	424
10 th graders	-.27	95	-.10	121	-.19	216
11 th graders	.11	45	.38	47	.24	92
12 th graders	.17	57	.31	73	.24	130
Emotional Impact						
8 th graders	-.32	37	-.32	12	-.32	49
9 th graders	-.26	197	.23	227	-.02	424
10 th graders	-.33	95	.01	121	-.16	216
11 th graders	-.16	45	.27	47	.06	92
12 th graders	.07	57	.48	73	.28	130
Parent Impact						
8 th graders	-.07	37	-.24	12	-.16	49
9 th graders	.02	197	.00	227	.01	424
10 th graders	-.06	95	-.17	121	-.11	216
11 th graders	.13	45	.07	47	.10	92
12 th graders	.08	57	.18	73	.13	130
Peer Impact						
8 th graders	.08	37	-.26	12	-.09	49
9 th graders	.00	197	-.01	227	-.01	424
10 th graders	.00	95	-.13	121	-.07	216
11 th graders	.01	45	.17	47	.09	92
12 th graders	.16	57	.03	73	.09	130

Table 5. *Gender And Grade Differences In The Impact On Students Who Were Told They Failed The Minnesota Basic Standards Test In Mathematics*

Source and Scale	df	Univariate F value	Significance	Partial η^2
Gender (main effect)				
Academic Impact	1	4.65	.03	.01
Extracurricular Impact	1	2.96	.09	.00
Emotional Impact	1	41.63	.00	.04
Parent Impact	1	.31	.58	.00
Peer Impact	1	.69	.41	.00
Grade (covariate)				
Academic Impact	1	24.21	.00	.03
Extracurricular Impact	1	8.74	.00	.01
Emotional Impact	1	9.16	.00	.01
Parent Impact	1	2.35	.13	.00
Peer Impact	1	1.11	.29	.00

Note. N = 911 (480 girls and 431 boys). We conducted a multivariate analysis of covariance (MANCOVA) testing gender as an independent variable and grade level (8-12) as a covariate, using standardized scores for each of the five impact scales. For gender, there was a statistically significant F approximation to Wilks's Lambda, $F(5, 904) = 12.595, p < .000, \eta^2 = .065$. For grade, the F approximation to Wilks's Lambda also was statistically significant, $F(5, 904), p < .000, \eta^2 = .029$.