

National Association of Test Directors 2001 Symposia

Organized by:

Mary E. Yakimowski-Srebnick, Ph.D.
**Hampton (VA) City Schools and Virginia Polytechnic Institute and State
University**

Edited by:

Joseph O'Reilly
Mesa (AZ) Schools

This is the seventeenth volume of the published symposia, papers and surveys of the National Association of Test Directors (NATD). This publication serves an essential mission of NATD - to promote discussion and debate on testing matters from both a theoretical and practical perspective. In the spirit of that mission, the views expressed in this volume are those of the authors and not NATD. The paper and discussant comments presented in this volume were presented at the April, 2000 meeting of the National Council on Measurement in Education (NCME) in New Orleans.

The authors, organizer and editor of this volume are:

Gregory J. Cizek
Associate Professor of Measurement and Evaluation
110 Peabody Hall
School of Education, CB 3500
University of North Carolina
Chapel Hill, NC 27599-3500

Mary Yakimowski

Michael Kean
CTB/McGraw-Hill
20 Ryan Ranch Road
Monterey, CA

Linda Fraser,
Robert Wetter,
Sharon Hutchinson
Kentucky Department of Education

Frankfort, KY

Joseph O'Reilly
Mesa Public Schools
549 North Stapley Drive
Mesa AZ 85203

TABLE OF CONTENTS

An Overview of Issues Concerning Cheating On Large Scale Tests	
Gregory J. Cizek.....	1
Security in a High-Stakes Environment: The Perceptions of Test Directors	
Mary E. Yakimowski-Srebnick.....	31
Cheating on Large scale Assessments: A Publisher’s Perspective	
Michael Kean.....	74
Kentucky Test Security	
Linda Fraser, Robert Wetter & Sharon Hendrickson.....	83
Discussant’s Comments	
Joe O’Reilly.....	102

An Overview of Issues Concerning Cheating on Large-Scale Tests

Gregory J. Cizek

University of North Carolina

Cheating undermines integrity and fairness at all levels. It leads to weak life performance. It undermines the merit basis of our society. Cheating is an issue that should concern every citizen of this country. (Cole, 1998, p. A-24)

Sound testing practices and the high quality information that can result, are helpful to those who have oversight, responsibility, or interest in American education. From a broader perspective, sound testing programs yield benefits to society at large (Mehrens & Cizek, 2001). To the extent that tests provide high quality information, better judgments about individual students can result. Test data also provide the grist for pursuing well-reasoned

courses of action in terms of recommendations for improving policies and practices and evaluating reforms.

Just as clearly, factors that attenuate the validity of tests or degrade the usefulness of information yielded by them represent threats to sound decision making. It can be said that the primary role played by those in the field of psychometrics is what might be called "data quality-control specialist"--helping to ensure that tests yield the kind of valid and useful information that they were designed to produce. One aspect of data quality-control is a professional vigilance about threats to the accuracy and dependability of test information.

To a great degree, modern testing theory and practice have evolved to potently address many of the threats. For example, validity theory has been advanced through the work of Kane (1992), Messick (1989), and others. Generalizability theory (Brennan, 1992) provides a sophisticated new way of examining the dependability of test scores. The state of the art in setting passing scores has advanced more in the past decade than perhaps over any other period (Cizek, 2001). Computerization has made automated test assembly and administration as common in high-stakes testing contexts as the #2 pencil (Luecht, 1998). The degree and breadth of these changes is witnessed by the recent, extensive revision to the *Standards for Educational and Psychological Testing* (AERA/APA/NCME, 1999).

Despite these advances, areas for increased vigilance remain. One perplexing challenge is presented by the problem of cheating on tests. As the use of tests for informing decisions increases, and as the stakes associated with test performance rise, the problem of cheating has become more prevalent. One need not consult psychometric journals for evidence of the phenomenon; even a casual reading of the local newspaper demonstrates that the incidence and extent of cheating is on the rise. A few examples:

* According to an Associated Press report, "At least 52 teachers from five states cheated on their competency tests by paying \$1000 bribes to exam supervisors for extra time and help with the answers" (Payne, 2000, p. A-7).

* A front-page article in the *Detroit News* reported on called for "a full investigation into the Dearborn [Michigan] Police Department's promotional tests, following a

prosecutor's report that concludes someone in the city altered test results" (Merx, 2000, p. A-1).

* Spanning just the last three years, stories in the *New York Times* have documented cheating on tests assessing a wide range of vocations, knowledge, and skills; tests of citizenship, truck driving, stock brokering, and others have made the headlines (see, Seelye, 1998; Toy, 1999; Sullivan, 1997). A prominent series of stories has chronicled the recent investigation of widespread cheating on student achievement tests by New York City educators (see Stancik, 1999).

SOME BACKGROUND ON CHEATING ON EXAMINATIONS

Although cheating on tests has increasingly been given attention in the popular media, it has received only modest attention by those actually involved in the field of testing. However, the topic is still one that is frequently swept under the carpet or ignored. To begin a consideration of cheating, it is appropriate to first define the kinds of testing situations in which cheating can occur. Then, in subsequent sections of this article, I will present a summary of some methods for detecting cheating and suggestions for how cheating can be prevented.

Contexts and Varieties of Cheating

In this article, I limit the domain of cheating to that which occurs on large-scale educational achievement tests such as the kinds of pupil proficiency testing increasingly mandated by individual states. These tests are typically administered to students at prescribed grade levels, (say, at grades 4, 8, and 12) to measure student progress at key junctures in their education (such as at exit from elementary, middle, and high school). Some states choose a off-the-shelf test such as the *Iowa Tests of Basic Skills* or *Stanford Achievement Test* for monitoring purposes. These tests are designed to assess a set of objectives that would be broadly common across the country. Other states contract with test development companies to produce custom tests that are carefully aligned to the state's unique curriculum framework.

Regardless of the source of the tests, the content covered by them is most often limited to basic subject areas such as reading, writing, mathematics, science, and civics. Usually, the format of the tests is a combination of the familiar multiple-choice questions and so-called

“constructed-response” items in which the student, for example, solves a mathematics problem, writes an essay, or conducts a science experiment. A pupil’s performance on the test may be reported as simply passing or failing, or a system of ordered descriptors of performance may be used, such as *Advanced*, *Proficient*, *Nearly Proficient*, and *Beginning*. Increasingly, consequences for students are being attached to performance on the tests. For example, in some states, passage of a 4th grade reading test may be required for promotion to the 5th grade, or passing the 12th grade test may be required in order to receive a high school diploma. Tests with weighty consequences such as these are referred to as “high-stakes” tests.

There are many forms of cheating that can occur on such tests. In general, cheating can be defined as any action that violates the rules for administering a test, any behavior that gives an examinee an unfair advantage over other examinees, or any action on the part of an examinee or test administrator that decreases the accuracy of the intended inferences arising from the examinee's test score or performance. A person may not actually take a test himself or herself, but may use another person (called a "confederate") to take the test in his or her place. An examinee may use unauthorized materials such as a "cheat sheet," or take advantage of the testing situation by, for example, requesting testing accommodations that are not necessary. One method of cheating on large-scale examinations involves taking advantage of time zone differences when tests are to be administered at sites across the U.S. at the same time. Using this method, examinees at one test site in, say, New York City, complete an examination in time to telephone other examinees in, for example, Los Angeles, to communicate the content of the test in advance. (To my knowledge, there are no organizations that address this method, despite the fact that it has been documented to have been used extensively to cheat on tests such as the Graduate Management Admissions Test (GMAT), the Test of English as a Foreign Language (ToEFL), and the Graduate Record Examination (GRE), and the fact that simply staggering test administration times would address the problem. Overall, these actions represent broad categories of behaviors; a more extensive and specific list of cheating methods is provided by Cizek (1999).

For their part, test developers usually produce carefully scripted directions for administering their tests and provide clear guidelines for what kinds of behaviors on the part of examinees are permissible and which are not. Acceptable and unacceptable behaviors are sometimes formalized in states' administrative codes or statutes. Numerous professional organizations

have published statements on the inappropriateness of cheating. Some of the most explicit statements regarding cheating are found in the aforementioned *Standards for Educational and Psychological Testing*. Among other things, the *Standards* indicate those involved in testing programs should:

- * protect the security of tests (Standard 11.7);
- * inform examinees that it is inappropriate for them to have someone else take the test for them, disclose secure test materials, or engage in any other form of cheating (Standard 8.7);
- * ensure that individuals who administer and score tests are proficient in administration procedures and understand the importance of adhering to directions provided by the test developer (Standard 13.10);
- * ensure that test preparation activities and materials provided to students will not adversely affect the validity of test score inferences (Standard 13.11); and
- * maintain the integrity of test results by eliminating practices designed to raise test scores without improving students' real knowledge, skills, or abilities in the area tested (Standard 15.9).

In summary, there has not been a dissemination problem regarding what constitutes integrity in testing, and what constitutes cheating on tests. Virtually everyone involved in testing knows how to administer (and take) tests that yield credible, accurate results. Unfortunately, mere knowledge about what constitutes cheating is not enough.

WHY CHEATING IS A PROBLEM

Validity is the single greatest concern in any testing situation. The concept refers to the accuracy of the interpretations about examinees based on their test scores. Phrased in only slightly more technical terms, validity is the degree to which evidence supports the inferences made about a person's knowledge, skill, or ability based on his or her observed performance. By definition, inferences are based upon a less-than-ideal amount of information, such as on a sample of a person's knowledge or skill obtained via a test. Because it is often too costly or impractical to gather more information, inferences must be based on samples of behavior. Consequently, it is necessary to consider the accuracy of inferences based on the available evidence (e.g., test performance); that is, to consider validity. This idea of validity as

accuracy-of-inferences and sufficiency-of-evidence are central in modern psychometric theory and are the foundation of professionally defensible testing practices. Any factor that attenuates the ability to make accurate inferences from the sample of performance threatens validity and jeopardizes the meaningfulness of conclusions about the test taker. When cheating occurs, inaccurate inferences result.

WHO CHEATS, HOW MUCH, AND WHY?

Test takers cheat. They let others cheat. Test administrators and proctors cheat. Although hard data on the frequency of cheating is difficult to come by, there is some evidence. The data we do have is of two types: results of research studies on cheating (most often surveys), and anecdotal reports that arise via newspaper and broadcast media outlets. Both sources of evidence have limitations. Surveys always suffer from some degree of inaccuracy; the concerns are heightened when the questions center on sensitive or illegal behaviors.

Anecdotal reports are sometimes exaggerated or prove to be false. Despite these limitations, the incidence of such anecdotes has increased dramatically, and there is enough credible research evidence accumulating to conclude that the problem of educators cheating on tests is occurring and is occurring more frequently. Summarizing across studies, a range of 3-5% is a reasonable estimate of the percentage of examinees who engage in cheating on any particular occasion. Anecdotal reports confirm that it is not only test takers who are cheating. As was documented in the cases described previously involving bribes paid to proctors and in the far-reaching investigation in New York City schools, those who give tests are also engaging in the behavior with surprising frequency.

It is easiest to comprehend examinees' motivations for cheating. They want a license to practice in their chosen profession, higher grades, opportunities for advancement, issuance of a credential, and so on. Sometimes examinees allow other test takers to cheat. One study by Davis, et al. (1992) conducted with college students, examined the reasons why they would do so, with the top reasons shown in Table 1.

Table 1

Top Reasons for Letting Other Students Copy During an Examination

8. Just to do it. I didn't like the teacher, and I knew if I got caught nothing would happen.
7. I knew they studied and knew the material, but test taking was really difficult.
6. No particular reason. It doesn't bother me because I probably got it wrong and so will they.
5. Because they might let me cheat off them sometime.
4. She was damn good looking.
3. I wouldn't want them to be mad at me.
2. I knew they needed to do good in order to pass the class. I felt sorry for them.
1. He was bigger than me.

Cheating by those who give tests is only slightly more difficult to understand. Basic motivations provided by bribes or other rewards certainly exist. Those who teach courses, direct residency programs, or oversee education and training organizations have an interest in promoting strong performance on the part of their students. Finally, numerous research studies have documented that the majority of high school and college graduates cheated on tests in their own academic careers; because so much cheating of that cheating went undetected and unpunished, and because they can easily put themselves in the position of examinees desperate to pass a test, those who give tests may often be tempted to turn a blind eye to cheating.

CHEATING AS AN IRREGULAR EVENT

A conclusion that cheating has occurred on a test can only be made after a careful examination of evidence. Usually, such an investigation begins following what is initially termed a "testing irregularity." When tests are administered, events that are out-of-the-ordinary can occur. These events may be within or beyond the control of those administering and/or those taking tests and, until causal attributions can be confidently asserted, cannot be interpreted as cheating. A first step in detecting cheating is to have in place a set of

procedures for observing and documenting irregularities. Examples of irregularities could include:

- 1) a fire alarm activation that required evacuation of a building during a testing session. Ordinarily, this event would be beyond the control of test administrators, but the event could increase student anxiety, reduce students' ability to attend to test materials on their return to the testing session, etc. If this occurred, students' performances on the test may not represent their true levels of knowledge, skill, or ability; that is, the students' proficiency levels would be underestimated.
- 2) permitting examinees to have additional time to complete a test beyond the limits indicated. This event would ordinarily be within the control of test administrators. If this occurred, examinees' performances on the test may again not represent their true levels of knowledge, skill, or ability, though in this case, students' proficiency levels would likely be overestimated.
- 3) repeated, sustained glancing by one examinee at the answer sheet of an adjacent examinee.

Two fundamental questions arise when a testing irregularities occurs. One question concerns the likelihood of the event. Unusual or out-of-the ordinary occurrences happen all the time; however, some events are less likely than others. The more unlikely an event is to occur, the more our curiosity is piqued. For very rare events, such as winning a Super Lottery or being struck by lightning, the probability of their occurrence is often of great interest.

The second question centers on explanations for unusual events. For example, airplane crashes are very rare; the intense interest in understanding the cause of that rare occurrence can linger for months, even years following the event. Our interest is particularly keen in understanding what role, if any, human intervention may have played in the event. Purely random events occur all of the time and they can be readily accepted as such. For example, in a fair lottery, numbers are selected randomly and those who do not hold the winning number can (usually) accept the randomness of that event. On the other hand, it would not be tolerable if human intervention or manipulation of the Lottery number selection tilted the process in favor of certain numbers or gave an *a priori* advantage to certain individuals. This type of human intervention changes our characterization (and acceptability) of the process from random to fraudulent.

The responsibilities of those who administer tests mentioned in the previous section are particularly germane to this point. When suspicion exists that testing irregularities may have

occurred as a result of human intervention--either through negligence, deviation from prescribed testing practices, or intentional manipulation of circumstances, testing conditions, or results--then our sense of ethical behavior and fairness is violated as are, in many cases, legal or administrative guidelines.

METHODS FOR EVALUATING TESTING IRREGULARITIES

There are two general categories of methods for investigating and evaluating testing irregularities: judgmental and statistical. As the label suggests, judgmental methods rely more heavily on subjective human judgments. For example, a student might enlist the aid of a confederate to take an examination in his or her place. Human judgment is involved in detecting and responding to this irregularity when the proctors for the examination scrutinize photo identification before permitting examinees to take the test. Judgment is also involved when handwriting samples from the student are compared with those of the confederate to make a determination of whose handwriting appears on the test materials.

Statistical methods can be used to estimate the likelihood of events, such as anomalous or unusual test results. Some events have probabilities associated with them that are very small. For example, the first-year National Hockey League team, the Columbus Blue Jackets, are estimated to have only a 1 in 500 chance ($p = .002$) of winning the 2001 Stanley Cup. Those odds are actually fairly good compared to chances of being struck by lightning are (1 in 709,260 or $p = .00000141$); the chances of dying from a lightning strike are even less (1 in 2,794,493 or $p = 000000358$). Worse yet are the chances of correctly picking six numbers out of 49 in a lottery (1 in 14,000,000 or $p = .000000071$).

All of the p -values mentioned in the preceding paragraph refer to extremely small probabilities. In fact, the examples selected illustrate occurrences that could be considered nearly impossible. What is the threshold that should cause us to consider an event as being so unlikely due to chance that we are compelled to consider other potential causes? In the social sciences, the standard probability level associated with statistical significance (that is, the p -

value at which scientists come to conclusions and/or make decisions about human behavior) is $p < .05$.

Of course, highly unlikely events *can* occur. However, as mentioned previously, we ordinarily become suspicious when highly improbable results occur, and we are led to conclude that simple chance should be ruled out as a plausible explanation. Should the Blue Jackets win the Stanley Cup, the circumstances surrounding such an upset in the expected course of events would likely lead to calls for an investigation of any irregularities in that sporting contest. Similarly, unusual results can occur on tests. For example, two examinees seated next to each other may and taking a 200-item multiple choice licensure examination, may answer 146 items correctly. Further, they may choose the same incorrect options for the 54 items they answered incorrectly. Statistical methods for detecting cheating on tests answer the simple question: “How likely is it that these examinees would, *by chance alone*, have produced the same response patterns?” If the answer to that question suggests that the events were *not* very likely due simply to chance, then investigations into plausible alternative explanations begins.

It is important to note, however, that statistical methods do not obviate the need for human judgment. Even once test results are shown to be highly unlikely, human rationality must be invoked to come to any conclusions about whether alternative causes represent more plausible explanations for the results; that is, there still exists a need to make subjective interpretations about whether the unlikely events represent cheating.

TRIGGERING INVESTIGATIONS OF TESTING IRREGULARITIES

It is not enough to ascertain that a testing irregularity was an improbable event. As mentioned previously, improbable events do occur. The probability of obtaining a score of 20 out of 20 through blind guessing on a test comprised of True/False items would be $p = .000000954$ --a nearly impossible event. However, other factors would ordinarily alter our interpretation of that probability. For example, if an examinee did not answer the 20 items through blind guessing, but used his or her knowledge of the content being tested to make more informed answer choices, then the probability would be substantially reduced. Further, if the test were

an extremely easy one, and if the examinee were highly knowledgeable, then the probability of obtaining a score of 20 out of 20 could approach $p = 1.0$. Thus, to evaluate the probability of an occurrence, we must bring ancillary information to bear.

One increasingly essential source of supplemental information is referred to as a “trigger.” In large testing programs such as the SAT, for example, many people obtain scores that are highly unusual (e.g., a total score of 1600). Such performance would not arouse suspicion of an irregularity if the student had taken the test previously and obtained a 1560, if the student had a high school GPA of 4.0, was class valedictorian at a college preparatory school, and so on. On the other hand, such performance *would* arouse suspicion, for example, if the examinee’s previous performance had been a 470, if a fellow student reported that the examinee had access to the SAT test questions in advance, or if a test proctor observed the examinee copying from a test-taker of extremely high ability who was seated nearby. Each of these latter situations involves what is called a trigger--additional information that suggests further investigation of the irregularity is warranted.

In cases cheating is suspected, statistical evaluations of test results are usually not appropriate in the absence of a trigger. However, the presence of a trigger necessarily changes our interpretations of the likelihood that results were obtained fairly. Suppose, for example, the 20-item true/false test described earlier involved simple multiplication facts. It would be highly unlikely for a three-year-old child to obtain a raw score of 20. Statistical estimates of the probability of the event would be very small, but the small probability would not necessarily lead to an allegation that the result was improper. However, if an observer of the child during the test reported that he or she saw the child’s parent whispering something in the child’s ear immediately prior to the child answering each of the questions, that information--a trigger--would suggest that the unusually unlikely event be regarded with a heightened level of suspicion, and that other, plausible explanations for the child’s amazing performance be investigated. Common triggers for conducting statistical investigations of alleged cheating include such things as observations by a proctor of unusual examinee behavior during an examination, or anonymous “tips” or reports that an examinee had access to a secure examination materials prior to the administration.

Of course, triggers usually involve human judgment and, as such, can be fallible. The extensive literature in the field of criminology speaks definitively about the unreliability of eyewitness testimony. An act of inference occurs when a proctor observes one examinee cheating by looking at another examinees' answer sheet. Objectively, the behavior can also be interpreted as an examinee innocently averting his or her gaze temporarily to gain relief from intense concentration on the task at hand.

STATISTICAL TOOLS

A number of statistical tools exist to help in the detection of possible cheating and provide quantification of the probability that an irregularity can be attributed to chance. One commercially-available software program has been developed, and is currently available from Assessment Systems Corporation (ASC) of St. Paul, Minnesota. The program is called *Scrutiny!* and it can be run on a typical personal computer. *Scrutiny!* uses an approach to identifying copying called "error similarity analysis" or ESA--a method which, unfortunately, has not received strong recommendation in the professional literature. One review (Frery, 1993) concluded that the ESA method: 1) fails to utilize information from correct response similarity; 2) fails to consider total test performance of examinees; and 3) does not take into account the attractiveness of wrong options selected in common. Bay (1995) found that ESA was the least effective index for detecting copying of the three methods she compared.

Despite this technical weakness, *Scrutiny!* and its accompanying documentation provide a thorough introduction to the logic of detection and sound advice regarding appropriate cautions for interpretation and use of the results. Additionally, the software is easy to use, is compatible with many common input file formats, permits the user to enter a seating plan (which can be used in conjunction with identified examinee seat locations to corroborate the suspicion of cheating) and produces traditional test summary statistics as well as the probability statistics of interest regarding potential answer copying.

Other statistical indices for detecting potential copying exist, though they are not yet available in commercially available software packages. Two such indices, g_2 , developed by Frery et al. (1977) and ω , developed by Wollack (1997) are technically superior to the ESA method.¹ These procedures offer more power to detect copying, while safeguarding against over-

identification of copying (i.e., Type I errors or false positives), and can be used with relatively small sample sizes (i.e., around 200 examinees). Unlike the ESA and other methods that rely only on common numbers of errors which can bias results when overall ability is not accounted for, g_2 and ω incorporate information from common right answers and differential probabilities of selection of incorrect options.

Although *Scrutiny!* or other methods may provide a defensible way of producing evidence to support a suspicion of cheating, it is important to restate that statistical analyses should be triggered by some other factor (e.g., observation). None of the statistical approaches should be used as a screening tool to mine data for possible anomalies. A recent court decision involving an the Association of Social Work Boards (ASWB) examination program provides an illustration. According to an article in the *ASWB Association News* (Atkinson, 2000), several examinees who had taken the February 1995 administration of the ASWB examination had their scores invalidated and were refused the issuance of licences. These actions were the result of analyses of their test scores which "revealed statistical abnormalities" (p. 9). In litigation, it was noted that "there did not appear to be any on-site problems" or reports of irregularities when the test was administered, although an "administrator for the social work board had received a telephone call indicating that certain individuals had copies of the exam prior to its administration" (p. 9). Both the circuit court and appeals court decided in favor of the examinees, noting that there was a lack of evidence to suggest why the examinees were investigated for possible cheating in the first place. It appears that, the telephone call notwithstanding, no triggering event was found such as would justify the consideration of statistical evidence.

THE PARTICULAR PROBLEM OF EDUCATOR CHEATING ON TESTS

The testing director of a large city school district summarized the problem: "Teachers cheat when they administer standardized tests to students. Not all teachers, not even very many of them; but enough to make cheating a major concern to all of us who use test data for decision making" (Ligon, 1985, p. 1).

One need only search the internet, look at an issue of a national magazine, or skim a newspaper, to confirm that many educators are attempting to circumvent the testing, monitoring, or accountability systems. Stories of cheating abound, and the methods are numerous, ranging from subtle coaching to overt manipulation. A *US News and World Report* article described a case in Ohio, where one educator is accused of physically moving a student's pencil-holding hand to the correct answer on a multiple-choice question (Kleiner, 2000). A recent *Washington Post* story announced the resignation of a Potomac, Maryland principal who stepped down amidst charges that she "was sitting in the [class]room, going through test booklets and calling students up to change or elaborate on answers" (Schulte, 2000). A colleague of mine in educational testing tells the story of how a principal would begin the announcements each morning with a greeting to students via the schools public address system: "Good morning students and salutations! Do you know what a salutation is? It means 'greeting,' like the greeting you see at the beginning of a letter." Apparently the students learned the meanings of words like "salutation" from the principal's daily announcements; they probably never learned that his choice of words like "salutation" wasn't done randomly, but was done with the vocabulary section of the state-mandated, norm-referenced test in hand.

I found out about one of the most blatant forms of educator cheating over a decade ago at an evening reception following a conference for school district superintendents in one midwestern state. I happened upon a conversation among several superintendents who, with cocktails in hand, were chuckling and winking about how their quality control procedures for student testing involved "pre-screening the kids' answer sheets for stray marks." What was so funny--I found out later from one of the superintendents--was that "stray marks includes things like wrong answers." Wink. Apparently, the practice continues. Another recent article describes how 11 school districts in Texas are being called to account for an unusually high number of erasures on that state's test (Johnston & Galley, 1999).

Most cheating is probably not as overt. More subtle forms of cheating are undoubtedly more frequent, but still serve to degrade the meaning of test results and confidence in education systems. The more subtle kinds of cheating occur when a teacher prods a student to review his or her answer: "Why don't you take another look at what you wrote down for number 17." Some of those who give tests cheat by proxy, when they fail to monitor test taking and

effectively encourage cheating on the part of students. Educator cheating also occurs when they fail to include all students who would be eligible to take a test, as is the case when a teacher reminds certain students who are likely to obtain low scores on a test that it would be OK for them to be absent on the day of the test. The *Education Week* article by Johnston & Galley (1999) described a sophisticated variation of this kind of cheating in which incorrect student identification numbers were apparently purposefully entered on the answer sheets of low-scoring students, which had the effect of kicking those answer sheets out of the scoring process and raising the school's average performance. In other states which require that an absent student be recorded a score of zero (which would lower a school's average performance) all students are encouraged to attend on the day of testing, but some are afforded "testing disability accommodations" such as an individual aid, reader, or other assistance not usually a part of the student's educational experience.

Perhaps the most visible report of cheating by educators involved teachers and principals in the New York City school system. An exhaustive study of cheating was conducted by Edward Stancik, Special Commissioner of Investigation for the New York City School District. The study found that cheating by 12 educators was "so egregious that their employment must be terminated and they should be barred from future work with the [Board of Education]" (Stancik & Brenner, 1999, p. 63) The report named another 40 educators who were recommended for disciplinary action 35 of whom engaged in actions judged serious enough to warrant potential termination. Examples of the cheating Stancik identified included those of a principal, who during a test "walked around the room and pointed out [to the students] incorrect choices, saying either "That's wrong" or "Do that one over" (p. 2). According to Stancik's investigation, 4th-grade students at another school reported that their teacher, Teresa Czarnowski, helped them cheat by correcting their answers in advance. Stancik reported: "According to one boy, who is indicative of those we interviewed, after he finished the test on the separate sheet [of scrap paper], he gave it to Czarnowski who checked his choices and marked an X on the scrap next to his wrong answers. Then she returned the paper to the student who corrected his responses and, finally, he transferred his selections to the official bubble form" (p. 11). Overall, the report concluded that there had been "extensive cheating by educators" that the school district had "known about the problem for years" and that "educators were no held fully liable for their misconduct" (p. 60). The public release of the

initial report brought greater attention to the problem. According to a follow-up report issued in May 2000 by the investigators' office:

“Almost immediately, our intake unit was busy with new complaints of wrongdoing committed by Board of Education employees during the testing process. Then in February 2000, while we were conducting investigations into those allegations, students took the State English Language Assessment (ELA) examination and reports of suspicious behavior and writing in test booklets again poured into our office.... Once again we found proctors who gave answers to students, alerted them to wrong responses, and changed student choices after the exam was turned in. Moreover, this investigation uncovered new methods of misconduct, including prepping children for the third day of the ELA exam by using the actual test material. Finally, our investigations continued to be impeded by delay in the reporting of testing allegations to this office.” (Stancik, 2000, p. 1)

The follow-up report named another 10 educators who had engaged in seriously inappropriate behaviors during testing in New York City. For many of the educators named, the cheating was so blatant--for example, writing answers to test questions on the chalk board--that immediate termination of employment was recommended.ⁱⁱ

RESEARCH ON EDUCATOR CHEATING

The most common avenue of research does not ask educators directly about whether they engage in what have come to be referred to euphemistically as “inappropriate test administration practices” though a few studies have done so. Usually, educators have been polled regarding their general perceptions of cheating in their schools. One such study asked 3rd, 6th, 8th, and 10th grade teachers in North Carolina to report how frequently they had witnessed certain inappropriate practices. Overall, 35% of the teachers said they had observed cheating, by either personally engaging in inappropriate practices or being aware of unethical actions of others. (The teachers in this study reported that their colleagues engaged in the behaviors from two to ten times more frequently than they had personally.) The behaviors included giving extra time on timed tests, changing students' answers on their answer sheets, suggesting answers to students, and directly teaching specific portions of a test. More flagrant examples included the case of students being given dictionaries and thesauruses by teachers for their use on a state mandated writing test. One teacher revealed that she checked students answer sheets "to be sure that her students answered as they had been taught." Other teachers reported more subtle strategies such as "a nod of approval, a smile, and calling attention to a

given answer" were effective at enhancing students' performance (Gay, 1990). Another study of teachers drawn from two large school districts found that 31.5% of the teachers surveyed reported spending two or more weeks giving students old forms of standardized tests for practice (Shepard & Dougherty, 1991).

In a study initiated to investigate suspected cheating in the Chicago Public Schools, a total of 40 schools were included, 17 as "control" schools and 23 "suspect" schools which exhibited irregularities in the performances of their 7th and 8th grade students on the *Iowa Tests of Basic Skills* (ITBS). The irregularities consisted of unusual patterns of score increases in previous years, unnecessarily large orders of blank answer sheets for the test, and high percentages of erasures on students' answer sheets. The researchers readministered the ITBS under more controlled conditions and found that, even accounting for the reduced level of motivation students would have had on the retesting, "clearly the suspect schools did much worse on the retest than the comparison schools" and concluded that "it's possible that we may have underestimated the extent of cheating at some schools" (Perlman, 1985, pp. 4-5). A study of cheating in the Memphis school district revealed extensive cheating on the *California Achievement Test*, including one case in which a teacher displayed correctly filled-in answer sheets on the walls of her classroom (Toch & Wagner, 1992).

Educators' Perceptions of Cheating

Perhaps the most troubling stream of research on cheating concerns the attitudes of educators toward cheating. Generally, there appears to be a growing indifference on the part of educators toward the behavior and even an increasing sense that cheating is a justifiable response to externally-mandated tests.

Several attempts have been made to investigate educators' perceptions of cheating. In one study, 74 pre-service teachers were asked to indicate how appropriate they believed certain behaviors to be. Only 1.4% thought that either changing answers on a student's answer sheet or giving hints or clues during testing were appropriate, and only 2.7% agreed that allowing more time than allotted for a test was acceptable. However, 8.1% thought that practicing on actual test items was okay, 23.4% believed rephrasing or rewording questions to be

acceptable, and 37.6% judged practice on an alternate test form to be appropriate (Kher-Durlabhji & Lacina-Gifford, 1992).

Table 2

Teacher beliefs about inappropriate test administration practices

Question: To what extent do you believe these are practiced by teachers in your school?

<u>Behavior</u>	Percent of respondents				
	<u>Never</u>	<u>Rarely</u>	<u>Often</u>	<u>Frequently</u>	<u>No Idea</u>
1. Providing hints on correct answers	28.5	20.8	16.9	5.8	28.0
2. Giving students more time than test directions permit	38.0	19.7	15.2	4.4	22.7
3. Reading questions to students that they are supposed to read themselves	38.8	22.2	11.9	2.2	24.9
4. Answering questions about test content	43.2	20.5	8.9	2.8	24.7
5. Changing answers on a student's answer sheet	58.4	7.8	5.5	0.6	27.7
6. Rephrasing questions during testing	36.3	20.8	16.1	1.9	24.9
7. Not administering the test to students who would have trouble with it	50.7	15.8	7.5	5.8	20.2
8. Encouraging students who would have trouble on the test to be absent on test day	60.1	10.8	5.5	1.9	21.6
9. Practicing items from the test itself	54.6	12.5	8.0	3.3	21.6
10. Giving students answers to test questions	56.8	11.6	6.4	1.9	23.3
11. Giving practice on highly similar passages as those in the test	24.9	15.8	20.5	19.7	19.1

From Shepard and Dougherty (1991)

The beliefs of pre-service teachers appear to translate into actual practices when they enter the classroom. A large sample of 3rd, 5th, and 6th grade teachers in two school districts was asked to describe the extent to which they believed specific cheating behaviors were practiced by teachers in their schools. On the positive side, their responses (shown in Table 2) indicated that for all of the behaviors listed but one, a majority of respondents said that they occurred rarely or never. Equally noticeable, however, is that a wide range of behaviors was reported as occurring “frequently” or “often” by, in some cases, 15% or more of respondents. A second observation that leaps from Table 2 is the remarkable extent to which teachers report that they have “no idea how often this occurs” (Shepard & Dougherty, 1991).

Another survey examined perceptions about two specific kinds of “test preparation” practices: having students practice for a state-mandated, norm-referenced test using another form of the

same test, or having students practice on the actual test that would be used. The survey polled six groups of educators, including teachers and administrators drawn from schools in the midwestern U.S., and teachers, principals, superintendents, and school board members from California. The results, shown in Table 3, reveal fairly broad acceptance of these behaviors, even among board members.

Table 3

**Teacher and administrator beliefs
about inappropriate test administration practices**

<u>Behavior</u>	Percent of respondents considering the practice to be appropriate					
	Midwest		California			
	Teachers	Administrators	Teachers	Principals	Supts.	Board Mbrs.
Student practice with previous test form	34	47	57	25	60	68
Student practice with current test form form	14	17	36	6	17	21

Though not attempted here (or elsewhere to my knowledge), the costs of cheating probably could be measured in dollars and cents. What cannot be measured are the effects of educator cheating at more fundamental levels. For example, when students learn that their teachers or principals cheat, what is the effect of this kind of role modeling? While fallen professional athletes might be able to say, “Don’t look at me as a role model, I am just an athlete doing a job,” educators cannot: a significant aspect of their job *is* the modeling of appropriate social and ethical behavior. Also, how might educator cheating affect students’ attitude toward tests or their motivation to excel? How might it affect their attitude toward education, their trust or cynicism with respect to other institutions, or their propensity to cheat in other contexts?

PREVENTING CHEATING

What can be done to deter cheating? Fortunately, many things. As a starting point, it is important to note that bringing the issue of cheating forward as a topic for discussion is likely to increase awareness of the problem on the part of those who give tests and those who take tests. It is important to heighten sensitivity about a validity threat heretofore virtually ignored. From the broadest perspective, it may be useful to entirely reconceptualize testing so that successful test performance can be more consistently and directly linked to student effort and effective instruction, and so that unsuccessful performance is accompanied by sufficient diagnostic information about students' strengths and weaknesses. As a result of identification and remediation of those weaknesses, we advance the perspective that obtaining accurate test results is more beneficial to all concerned than cheating (Cizek, 1999, chap. 11). There are also numerous, more pragmatic steps that can be taken. The following list should provide a start. Of the following, some are focussed on test givers; some on test takers; some apply to both.

- 1) Get the word out. It has been said that we more often stand in need of being reminded than we do of education. As mentioned previously, nearly all testing programs provide documentation describing appropriate test administration procedures, state regulations define legal conduct for test administrators, and professional associations have produced documents to guide sound testing practice. Nonetheless, reports of those accused of cheating on tests are often accompanied by the protestation that they did not know the behavior was wrong. If only as a reminder and to heighten awareness, every implementation of high-stakes tests should be accompanied by dissemination of clear guidelines regarding permissible and impermissible behaviors. Such reminders should be clearly-worded, pilot-tested, distributed, and signed by all who handle testing materials, including test site supervisors, proctors, and examinees.

- 2) Decrease reliance on easily-corruptible test formats. Changes in test development practice can reduce the potential for some methods of cheating. For instance, it is more difficult for one examinee to copy another examinee's answer to an essay question, case analysis, or other constructed-response format than it is to copy a

bubbled-in response or provide the key to a multiple-choice item. It must be recognized, however, that such changes require tradeoffs in terms of efficiency and scoring costs.

3) Limit the amount of testing. It is probably a truism that limiting the amount of testing will decrease the amount of cheating. As many states continue to expand their pupil proficiency testing programs as a primary mechanism for accountability, opportunities for cheating are expanded. There have been two, common, reactionary responses to the predictable increase in cheating. One reaction is the demand that large-scale testing for accountability be abandoned. For example, the September 22, 1000 issue of the *Congressional Quarterly* contained an essay by Monte Neill, the executive director of a group critical of testing, who argued the "pro" position on the question "Should high-stakes tests be abolished in order to reduce cheating?" (Neill, 2000). In the same issue, education writer Alfie Kohn is noted as one of several critics who "have seized on cheating as just another in a long list of reasons to abandon [standardized] tests." According to Kohn, "The real cheating going on in education reform is by those who are cheating students out of an education by turning schools into giant test-prep centers" (quoted in Koch, 2000, p. 759).

Related to the first reaction are demands that responsibility for judging student achievement be located more within individual educators' sphere of professional responsibility. A second reaction is that testing for accountability rely more heavily on constructed-response type item formats which, ostensibly, would be less prone to corruption. For instance, it is argued that it is more difficult to forge or coach a student's answer to an essay question or a science experiment than to alter a bubbled-in response or provide the key to a multiple-choice item.

The difficulty with these first-blush reactions is that they fail to fully address the core issues. As I have argued elsewhere, the genesis of high-stakes pupil testing in the 1970s was made inevitable because of poor decision making--or at least perceived poor decision making--and the resulting search for alternatives (identifying reference omitted). It was during the tumultuous 1970s that complaints of some business and industry leaders began to receive broad public currency: "We are getting high school

graduates who have a diploma, but can't read or write!" As Popham observed at the time: "Minimum competency testing programs ... have been installed in so many states as a way of halting what is perceived as a continuing devaluation of the high school diploma" (1978, p. 297). The clear public perception was that the gatekeepers were leaving the gates wide open. Perhaps a widespread misunderstanding of the relationship between self-esteem and achievement was to blame.

Understandably, educators wanted all students to achieve and all to have the personal esteem associated with those accomplishments. But assigning higher grades to heighten self-esteem and stimulate accomplishment too often had neither effect. The sense that grades weren't all they were cracked up to be wound its way from business and industry leaders' lips to policy makers' pens.

As the line of reasoning went, if the gatekeepers of the 1970s weren't keeping the gates as conscientiously as the public had hoped, then important decisions about students should be remanded to rely on passing one or more common tests. Thus, the obvious error in current calls to return to the past is that such a strategy only puts American education back in a place that caused accountability tests to be introduced in the first place. Moreover, though current tests have been shown to be susceptible to cheating, the solution of returning to measures and procedures that were demonstrably even more easily manipulated is unthinkable.

What should be considered is limiting the amount of testing for accountability. We must remember that there is a distinction between instruction and evaluation. It is obvious that not all tests are done for the purposes of evaluation. Equally true, however, is that not all tests--especially those designed for purposes of decision making--must have instructional value. Once their purpose has been clarified, the scope of mandated accountability tests, the time required for their administration, and the opportunities for cheating can be minimized.

4) Revise "Truth in Testing" laws. States with so-called "truth in testing" laws should reconsider the relative benefits such laws. These laws often require that the questions on tests used to monitor student achievement or for accountability purposes

be publicly disclosed following administration of a test. Despite their good intention, the unforeseen consequence of such laws has been an increase in educators' use of previous versions of tests for classroom practice, resulting in further narrowing of instruction. Additionally, the economic costs to states with such laws has been staggering, brought about by the need to develop entirely new monitoring instruments one or more times each year.

5) Audit test security procedures. Those with oversight for testing programs can incorporate operational changes--many of which require only modest changes in current procedures--that can have a cumulative positive effect on reducing cheating. Many of these are not new, and many may already be in place. However, a regular "security audit" to review procedures is desirable. Common security measures include shrink-wrapping, numbering, and bar-coding of test materials to deter unauthorized access and to permit tracing the path that the materials take. Other simple steps can easily be added, such as delaying delivery of testing materials until just prior to test administration, and, once delivered, requiring that materials be maintained securely by a named person responsible for the materials.

6) Improve test administration conditions. Increased attention must be paid to one of the weakest links in the security chain: proctoring. Too often, the qualifications for supervising or proctoring examinations are only faintly spelled out, the training provided is minimal if any, and no incentives exist to heighten their vigilance or pursue instances of cheating. For all testing contexts, proper training must include instruction on methods examinees use to cheat, how to approach a test taker regarding suspicions of inappropriate behavior without unduly disrupting other examinees or inducing anxiety in those who are not cheating. In the context of large-scale testing, training should include effective procedures for documenting on-site testing irregularities.

7) Use available statistical tools. Finally, we recall that statistical detection methods should not be used as screens for statistically unusual response patterns. Nonetheless, research has demonstrated that informing examinees that detection software will be used can dramatically reduce the incidence of cheating. One study by Bellazza and

Bellazza (1989) showed a reduction from approximately 5% to 1% in the amount of cheating on college-level management course examinations. If a detection program may be used to provide supplemental evidence following a triggering event, it makes sense to inform examinees that detection software may be used.

8) Provide penalties for cheating and change the system of investigation. In conjunction with limiting opportunities for cheating, procedures for investigating cheating and penalties for educator cheating must be dramatically revised. Currently, many tests are administered behind closed classroom door with little independent oversight; there are strong disincentives for educational personnel to report cheating; and in most jurisdictions, the responsibility for investigating cheating involves personnel at the school or district level and agencies such as boards of education with an inherent conflict of interest when it comes to ferreting out inappropriately high, apparent, student achievement.

Revised procedures should include: random sampling and oversight of test sites; increased protections for “whistle-blowers;” more streamlined procedures and stiffer penalties including permanent disqualification from teaching within a state and more coordinated sharing of information regarding educators who have had their licenses revoked; and delegation of responsibility for investigation cheating to an independent authority.

9) Implement honor codes. Because honor codes have been shown to reduce the incidence of cheating in other contexts, their use in licensure and certification testing should be examined. Honor codes require examinees to pledge to abide by a set of standards, including eschewing cheating themselves and obligating themselves to report cheating by others. Requiring examinees to sign such a pledge prior to taking an examination may work in credentialing settings as well.

CONCLUSIONS

Overall, the evidence on the problem of cheating on tests is in. Cheating is occurring with increasing frequency. It is fair to conclude that the problem will not disappear. The problem must be addressed, however, in order to ensure the integrity, fairness, and validity of test results. As a beginning step, those who have oversight of testing programs should make themselves aware of the myriad ways cheating can occur, including cheating by examinees and test administration staff who may aid examinees in cheating. Additionally, those responsible for testing programs should address how they can help to reduce cheating, and should pursue courses that foster even greater levels of public protection and professional responsibility for the citizens and associations they serve.

REFERENCES

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Atkinson, D. (2000, August). Testimony tests test. *ASWB Association News*, p. 9, 11.
- Bay, M. L. G. (1995, April). *Detection of cheating on multiple-choice examinations*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Bellazza, F. S., & Bellazza, S. F. (1989). Detection of cheating on multiple-choice tests by using error-similarity analysis. *Teaching of Psychology*, 16(3), 151-155.
- Brennan, R. L. (1992). Generalizability theory [NCME instructional module]. *Educational Measurement: Issues and Practice*, 11(4), 27-34.
- Cizek, G. J. (1999). *Cheating on tests: How to do it, detect it, and prevent it*. Mahwah, NJ: Lawrence Erlbaum Associates.
- 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* (pp. 3-17). Mahwah, NJ: Lawrence Erlbaum.
- Cole, N. (1998, November 9). Teen cheating hurts all. *USA Today*, p. A-24.

Davis, S. F., Grover, C. A., Becker, A. H., & McGregor, L. N. (1992). Academic dishonesty: Prevalence, determinants, techniques, and punishments. *Teaching of Psychology, 19*(1), 16-20.

Frary, R. B. (1993). Statistical detection of multiple-choice answer copying: Review and commentary. *Applied Measurement in Education, 6*(2), 153-165.

Frary, R. B., Tideman, T. N., & Watts, T. M. (1977). Indices of cheating on multiple-choice tests. *Journal of Educational Statistics, 2*, 235-256.

Gay, G. H. (1990). Standardized tests: Irregularities in administering of tests affect test results. *Journal of Instructional Psychology, 17*(2), 93-103.

Johnston, R. C., & Galley, M. (1999, April 14). Austin district charged with test tampering. *Education Week*, p. 3.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*, 527-535.

Kher-Durlabhji, N., & Lacina-Gifford, L. J. (1992, April). Quest for test success: Preservice teachers' views of high stakes tests. Paper presented at the annual meeting of the Mid-South Educational Research Association, Knoxville, TN. (ERIC Document Reproduction Service No. ED 353 338)

Kleiner, C. (2000, June 12). Test case: Now the principal's cheating. *U. S. News and World Report*.

Koch, K. (2000). Cheating in schools. *Congressional Quarterly, 10*(32), p. 759.

Ligon, G. (1985, March), Opportunity knocked out: Reducing cheating by teachers on student tests. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 263 181).

Luecht, R. M. (1998). Testing and measurement issues: Automated test assembly in the era of computerized testing. *CLEAR Exam Review, 9*(2), 19-22.

Mehrens, W. A., & Cizek, G. J. (2001). Standard setting and the public good: Benefits accrued and anticipated. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 477-485). Mahwah, NJ: Lawrence Erlbaum.

Merx, K. (2000, August 11). Cop test altered in Dearborn. *The Detroit News*, p. A-1.

Messick, S. A. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement, third edition* (pp. 13-104). New York: Macmillan.

Neill, M. (2000). Should high-stakes tests be abolished in order to reduce cheating? *Congressional Quarterly, 10*(32), p. 761.

Payne, P. (2000, August 18). Officials say 52 teachers paid \$1,000 to pass competency tests. *The [Schenectady, NY] Daily Gazette*, p. A-7.

Perlman, C. L. (1985, March). Results of a citywide testing program audit in Chicago. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 263 212), pp. 4-5.

Popham, W. J. (1978). As always, provocative. *Journal of Educational Measurement*, 15, 297-300.

Popham, W. J. (1991, April). Defensible/indefensible instructional preparation for high-stakes achievement tests. Presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Schulte, B. (2000, June 1). School allegedly cheated on tests. *Washington Post*, p. A-1.

Seelye, K. Q. (1998, January 28). 20 charged with helping 13,000 cheat on test for citizenship. *New York Times*, p. A-1.

Shepard, L. A., & Dougherty, K. C. (1991). Effects of high-stakes testing on instruction. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 337 468)

Stancik, E. F. (2000, May 2). Correspondence to Harold O. Levy, Chancellor of New York City Public Schools, pp. 1-2.

Stancik, E. F., & Brenner, R. M. (1999). *Cheating the children: Educator misconduct on standardized tests*. New York: Office of the Special Commissioner of Investigation for the New York City School District.

Sullivan, J. (1997, January 9). 53 charged in brokers' testing fraud. *New York Times*, p. A-7.

Thacher Associates. (2000). *Setting the record straight: Anatomy of a failed investigation*. New York: Author.

Toch, T., & Wagner, B. (1992, April 27). Schools for scandal. *U.S. News and World Report*, pp. 66-72.

Toy, V. S. (1999, November 23). Drivers' test scheme reveals secret decoder watchbands. *New York Times*, p. B-2.

Wollack, J. A. (1997). A nominal response model approach for detecting answer copying. *Applied Psychological Measurement*, 21, 307-320.

Security in a High-Stakes Environment: The Perceptions of Test Directors

Mary E. Yakimowski-Srebnick, Ph.D. ¹

Hampton (VA) City Schools and Virginia Polytechnic Institute
and State University

The use of student performance on tests in accountability systems is not new What is somewhat different about the current emphasis on performance-based accountability is its pervasiveness Student achievement is being used not only to single out schools that require special assistance, but also to provide cash incentives for improvements in performance. Yet several fundamental questions remain about the student assessments, the accountability model, and the validity, impact, and credibility of the system. (Linn, 2000, pp. 1-2)

Most Americans favor student testing for the purposes of information, accountability, and incentives (Phelps, 1999). Linn (2000 a, b) contends that this focus upon student testing has played a prominent role in many education reforms. The recent rise in the number of tests administered in the United States, especially those with consequences, has been said to coincide with the standards reform movement of the 1990s (Elmore, Abelman & Fuhrman, 1996).

HISTORICAL PERSPECTIVE ON TESTING

Testing and accountability have played “prominent roles in many of the education reform efforts during the past 50 years” (Linn, 2000 b). In the 1950s, testing was used routinely to select students to enter special programs for the gifted within our K-12 schools and for our institutions of higher education. Mehrens and Lehman (1987) contend that there was a significant shift that occurred between the early 1960s, a time from when many criticized tests, and that of the 1970s when the majority of individuals tended to favor testing.

The 1980s were a time of concern from special interest groups. This time period resulted in many truth-in-testing bills that generally required test developers to share information such as how the test was developed, known reliability and validity properties, appropriate uses of results, and confidentiality regulations. During this same time a number of legislative bills were enacted to implement statewide testing programs. Initially, these statewide programs focused on implementing competency tests designed to ensure that high school graduates had obtained a minimum level of

basic skills, generally in reading and mathematics. However, by the late 1980s and 1990s there was a public outcry that our graduates were not prepared to enter the workforce in the new millennium. This outcry gave rise to what has been called the Standards Movement, resulting in higher outcome measures established for our schools. These outcome measures, generally operationalized as higher test score performance, sought to provide a well-educated American workforce in the global economy (Gordon, 2001).

Phelps (1999) states that over 40 million tests are administered annually. Studies reporting the costs associated with administering these tests indicate estimates ranging from about \$575 per student (Haney, Madaus, & Lyons, 1993) to between \$848-1,792 per student (Picus & Tralli, 1998). Most recently, Hoff (2001) and Danitz (2001) indicate that results from a survey of state officials revealed a nationwide annual total of \$400 million with California testing programs in the lead having a \$44 million price tag. The general public, even across different respondent groups, vehemently favors more testing and higher stakes. According to Phelps (1999), "Twenty-seven polls taken between 1970 and the present asked specific respondents whether they thought education would improve if there were higher stakes in school testing. In 26 of the 27 polls, the answer was yes, in most cases by huge margins" (p. 15). One may ponder whether those responding to these polls are aware of the costs.

Evolution and Impact of High-Stakes Testing

According to Wiersma and Jurs (1990), “Some tests ... get more publicity ... (and) are acclaimed as important indicators of educational performance, or they are maligned as unfair tools of the educational system that have serious negative consequences for the examinees. Unfortunately, the proponents and critics are often talking about the same test. The tests are controversial because there are important consequences that are attached to high or low scores ... these are sometimes referred to as high-stakes tests” (p. 353).

Many (*e.g.*, Corbett & Wilson, 1991; Phelps, 1999; Phelps, 2000; Ravitch, 1995; Resnick & Resnick, 1985; Schwatz & Viator, 1990) have debated the various benefits and disadvantages of the use of high-stakes tests as a means of accountability and levers for change in support of higher educational standards. Advocates generally cite enhancement of student performance outcomes and equity. Phelps (1999) has tried to summarize some of the disadvantages. He indicates that the anti-testing “canon includes allegations that standardized tests, particularly those with high-stakes:

- ◆ Induce ‘teaching to the test’ which, in turn, leads to artificial inflation of scores;
- ◆ Narrow the curriculum to a small domain of topics;
- ◆ Tap only ‘lower-order thinking’ and hence discourage innovative curricula and teaching strategies;
- ◆ Cause student achievement to decline;
- ◆ Are unfair to minorities and women;
- ◆ Are costly in terms of money and time;
- ◆ Are overused in the United States, especially in comparison with other countries; and
- ◆ Are opposed by all those who truly care about children” (pp. 3-4).

PREVALENCE OF CHEATING/SECURITY BREACHES ON HIGH-STAKES TESTING

According to Canner (1992) and NCME (1991), high-stakes testing has increased the pressure placed on educators to improve test scores in both appropriate and inappropriate ways. A review of recent studies (*e.g.*, Colton, 1997; Grant, 2000; Hall & Kleine, 1992; Kher-Durlabhji, Neelam & Lacina-Gifford, 1992; Linn, 2000 a, b; Mehrens, 1998; Monsaas & Englehard, 1991; Perlman, 2000; Shepard, 2000; Shepard & Dougherty, 1991; Smith, 1991; Wellhousen & Martin; 1991) and a plethora of newspaper articles appears to indicate that the incidences of test security breaches leading to cheating on standardized tests, in general, and those with high-stakes, in particular, are extensive.

In a national survey of 356 high school teachers conducted by The American School Board Journal (Bushweller, 1999), 90% said cheating, though not just on standardized tests, is a problem in schools. Half the teachers said they encounter cheating by students in most of their classes; but cheating is not reserved just for students. In surveying 2,567 teachers, testing coordinators, principals and superintendents, Hall and Kleine (1991) found that 44% of the survey respondents reported that their colleagues engage in practices that are blatant cheating.

Similarly, Monsaas and Englehard (1991), using Fishbien and Ajzen's theory of personal action of the relationship of attitudes and subjective norm (*i.e.*, perceived

pressure) accurately predicted teachers' testing behavior. Their research concluded that there was a relationship between teacher attitudes and behaviors. As the amount of perceived pressure increased, justifications to cheat became more prevalent. Grant's (2000) qualitative study further explored the extent standardized testing influenced teachers. Grant found that teachers were not opposed to change and accountability, but sought professional development to learn about change and the rationales and consequences of such testing. Kher-Durlabhji, Neelam and Lacina-Gifford's study in 1992 examined the pre-service perceptions of teachers in score-enhancing techniques such as checking completed answer sheets, teaching test-taker skills, and preparing learner objectives based on standardized test items. Additionally, Wellhousen and Martin (1995) found that over half of the 63 teachers in their study indicated that they would cheat under certain conditions. Reasons included perceived benefits to the students or the deemed inappropriateness of the administered test. Assistance (*i.e.*, cheating) could include hints, rewording items, and teaching to the exact test.

Issues related to inclusion (*i.e.*, participation rates) and test preparation may also lead to cheating. Some researchers (*e.g.*, Linn, 2000 a, b) have postulated that it is necessary to safeguard against selective exclusion of students. Other researchers (*e.g.*, Shepard, 2000; Smith, 1991) have bemoaned that high-stakes tests cause teaching to the test. For instance, Smith (1991) studied test coaching by surveying Arizona teachers and administrators. This resulted in the creation of a typology of teacher orientation toward preparing students for high-stakes testing. Individuals such as Ligon (1995, 2000), O'Reilly (ND), Perlman (2000), and Protheroe (2000) have

written about appropriate vs. inappropriate test preparation activities. Appropriate deemed to be appropriate are used to help students score well without resorting to things such as test-score inflation, curriculum distortion, lower skills, and equity issues. Many of these individuals also point out that the nature of the tests, for example, criterion-referenced versus norm-referenced testing, need to be considered. Colton (1997) further postulates that even the new advancement towards computer testing is at least as vulnerable as paper/pencil.

The media has also reported numerous allegations of cheating. Andrews (2001) reported on Alabama teachers penalized for compromising test security, while Hopkins (1999) reported the allegation of a past employee of a State Education Administration copying and distributing portions of the mathematics test. In other instances, Orlans (2000) covered the story of principals and teachers in 32 New York schools providing questions in advance to students for them to practice and Borja (1999) reported the allegation of a middle school administrator in Virginia copying questions from the Standards of Learning exam and then distributing comparable ones to teachers and students. Bradley (1999) published an article on how entire sections of a test were published in a newspaper run by teachers while Webster (2000) has shared that Nevada high school students passed around answers during administration of the state proficiency examination. From these and numerous other recent media accounts, it appears that neither the students nor educators in any part of the United States are untainted from this allegation of cheating resulted from breaches in test security.

REASONS FOR CHEATING AND STEPS TO CONSIDER TO DETER IT

While various researchers and individuals representing the media have excoriated cheating, Bushweller (1999) identifies potential reasons for test security breaches. He postulates that erosion of ethics in self-centered culture, habits gained in cooperative learning groups, and teachers who do not wish the hassle of disciplining cheaters are underlying factors contributing to cheating problem.

O'Reilly (ND) has provided us with a model of what is needed to actually obtain a student's true test score (See Figure 1). Cheating and other security breaches would be considered pernicious to one's ability to obtain an accurate test score.

$$\text{Accurate Test Scores} = \text{Content Knowledge} + \text{Familiarity} + \text{Motivation}$$

Figure 1. O'Reilly's model of factors contributing to an accurate test score.

Phelps (1999) has indicated that "The critics unfairly compare high-stakes standardized testing to their own notion of perfection. Administration of high-stakes tests will never be perfect. There will always be some teachers and pupils who cheat. There will always be some students who are better prepared to take a test than others ... " (p. 17).

Others such as state department officials have taken a more positive approach as found by Mehrens' study survey of state education officials on test security practices.

Additionally, individuals such as Bushweller (1997), O'Reilly (ND), Popham (2000

b), and Kilian (1992) feel that professional development will help deter cheating. They address the importance of providing professional development to help ensure that individuals understand proper test administration, preparation practices, and other appropriate factors in ascertaining a student's accurate test score.

Additionally, others (*e.g.*, Colton 1997) have suggested we integrate technology to be used as countermeasures by proctors and staff trained to enhance effective observation skills and control the exposure to the examination.

To deter cheating, we have also seen the emergence and/or study of standards (*e.g.*, AERA, APA, NCME, 1985, 1999), guidelines (*e.g.*, NCME, 1991; Kimmel 1997; Wilson, 1993) and position statements (*e.g.*, AERA, 2000) tied to issues such as the appropriate use of tests and test security. The severity of this issue is evidenced in Wilson's (1993) writings where he strongly recommends that School Boards adopt codes of ethics for standardized tests that result in the dismissal of teachers for breaching security.

PURPOSE AND SIGNIFICANCE OF THIS STUDY

Though many stakeholder groups have been involved in studies (*e.g.*, Hall & Kleine, 1992; Mehrens, 1993) seeking perceptions about the importance and impact of tests, others (*e.g.*, AERA *et al.*, 1999) have advised and/or worked to implement various guidelines and standards to enhance test security. Yet, a review of the literature located no study that has been conducted focusing on exploring district test directors'

perceptions on the issue of enhancing test security. Thus, the purpose of this exploratory study was to focus on the general research question of describing district level test directors' perceptions about test security in a high-stakes environment. This included perceptions on the (1) prevalence of identified standards, (2) importance and existence of actions, procedures, policies and (3) consequences of high-stakes testing. An undergirding purpose of this study was to gather from test directors actual vignettes of student and educator breaches, and the hope to add to the literature on how to deploy solutions to help deter cheating and enhance test security.

METHODOLOGY

Sample

I drew the sample for this study from the membership of the National Association of Test Directors (NATD), an organization of professionals with responsibilities for administering assessment programs in K-12 public educational settings. This membership is broadly representative of North America. Members are generally employed by local or state education agencies, test publishers, or universities. Using the NATD mailing list, I selected the members who worked for local education agencies, herein referred to as school districts. Reasonable attempts were then made to ensure that only one individual from a district received a survey.

Instrumentation

The cover letter accompanying the mailed two-sheet questionnaire included the purpose of the canvass and the reason the individuals were selected to participate in the study. (See Appendices A and B.) Three follow-up e-mails were forwarded to

NATD members who were enrolled on the listserv. A moderate percentage of the sample selected to receive the survey was assumed to be enrolled on the listserv. The listserv was used as a vehicle to announce the study, allow individuals an optional opportunity to receive/return surveys electronically, and provide a professional reminder to submit completed surveys.

The first of four sections on the survey asked questions related to district and director characteristics. This included the type of district, state in which the district was located, student enrollment, and grade configuration. The percentages of students in the respective district eligible for free/reduced lunch, English as a Second Language (ESL)/Limited English Proficient (LEP), and special education were requested in order to determine representativeness and generalizability of the findings. Within this first section the respondents were asked four questions dealing with the level of experience these administrators had with the issue of test security. More specifically, respondents were asked to indicate the number of years in which they served as a testing director and the number of years until retirement. Questions were posed as to whether the district implemented a high-stakes testing program and whether there had been a test security breach within the district in the past two years.

The second section of the survey listed statements from the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1985), hereafter referred to as *The Standards*. *The Standards* document provides “the basis for evaluating the quality of testing practices as they affect the various parties

involved” (p. 1). The survey asked participants to respond to seven statements using a five-point Likert-type scale to indicate how often each activity, or standard, occurred in their district (“5” indicated “always”, a “1” meant “never”). The seven standards cited on the survey were noted by Cizek (1999) as primary standards related to cheating and test security issues. It is of special note that primary standards are “those that should be met by all tests before their operational use and in all test uses, unless a sound professional reason is available to show why it is not necessary, or technically feasible, to do so in a particular case” (AERA, APA, NCME, p.2).²

The third section included a series of 24 statements based upon the review of literature (e.g., NCME, 1991) and assistance from 14 educational administrators from urban, suburban and rural communities. These statements listed actions, procedures and/or policies often mentioned as practices to help maintain test security in order to deter cheating. For each of these statements, individuals were asked to respond to two questions. First, the director was asked how important the practice was to maintaining test security. Second, the director was asked to indicate the degree to which the practice existed in the school district. A five-point Likert response format was used for each question, with the higher score meaning higher importance or greater existence. A “difference” score was then calculated between perceived importance and existence.

² While an updated version of *The Standards* has been released (AERA, APA, NCME, 1999), it was felt by the researcher that there were minimum changes made with this second edition on the identified standards and that most people who knew about the document had at least a working knowledge of the earlier edition.

The fourth section was designed to obtain test directors' perceptions on consequences of high-stakes testing and overall direction of change. Two items related to perceptions of the degree that high-stakes testing enhanced student performance outcomes and school improvement efforts. Individuals were asked the degree to which there were higher stakes attached to test results over the past five years and whether they felt this would be increasing in the next five years. They were also asked whether they felt that high-stakes testing increases the potential for breaches in test security. A question was then presented to determine whether test directors felt that test security in a "high-stakes" environment could be a reality.

The survey concluded with two open-ended items. With confidentiality assured, individuals were asked to briefly describe a student-related and/or educator-related test security breach that had occurred in their district.

Data was entered into the Statistical Package for the Social Sciences (SPSS, 2001) with appropriate measures of central tendency and dispersion calculated. Bivariate correlations also run to explore the relationship between reported existence of a security breach and size of district and perceived existence of high-stakes testing program. Correlated t-test analyses were conducted to explore statistically significant differences between perceived importance and existence on the 24 statements tied to actions, procedures, and/or policies implemented to maintain test security.³ Open-

³ While initially an exploratory factor analysis with a Varimax rotation was planned to determine underlining domains of actions, procedures and/or policies implemented to address security, this analysis was conducted though not reported within this paper due to the limited sample size. It had

ended survey responses were interpreted using a causal-comparative thematic analytical approach.

RESULTS

Nonresponse Bias

A total of 66 surveys were returned. Response profiles were analyzed to provide some insight regarding the important issue of nonresponse bias (Kerlinger, 1986). Findings indicated that the respondents appeared to reflect a reasonable distribution of NATD members from school districts in the sample and were reasonably representative of these characteristics. No contention is made, however, that members of NATD are necessarily representative of school districts across the United States.

District Characteristics

The 66 respondents represented school districts with enrollments totaling 4,425,191 students. District enrollments ranged from 2,680 to 1.1 million students with a median of 25,011 students (see Table 1). Though the vast majority of the school districts (97.0%) had a PreK-12 or K-12 configuration, one respondent indicated a K-6 and another a K-8 enrollment configuration. The primary classification type for 33 (50.0%) of the districts was urban, compared to 20 (30.3%) suburban, and 4 (6.1%) rural. The analysis of the percentage of students on free/reduced lunch, English-as-a-

earlier been hoped that the factor analysis would aid in exploring the relationship of each factor with whether a district had experienced a security breach in the past two-years and whether a high-stakes

Second Language (ESL) or Limited English Proficient, and special education revealed variations among respondents. The least amount of variation was evidenced in the percentage identified as special education (*i.e.*, 5 to 22%).

Test Director Characteristics

A wide range in years of experience as a test director also was indicated ranging from two months to 25 years, and anywhere from an anticipated retirement this year to 30 years from now. While 81.8% of the respondents indicated that their district was implementing a high-stakes testing program, 42.4% noted that a breach in test security had occurred in the past two years.

Table 1

Descriptive Statistics of Student, District and Personal Characteristics Represented by Survey Respondents

	<u>Mean</u>	<u>Mdn</u>	<u>SD</u>	<u>Range</u>
Student enrollment	67,048	25,011	166,753	2,680 - 1,097,320
Characteristics of students				
Percent on free/reduced lunch *	39.6	35.0	21.2	0 - 85
Percent ESL/LEP *	11.0	4.8	12.1	0 - 44
Percent special education *	12.8	12.0	3.6	5 - 22
Characteristics of directors				
Number of years as testing director	9.0	7.5	6.8	2 mo. - 25 yrs.
Number of years until retirement	8.7	7.0	7.7	0 - 30
<hr/>				
		<u>n</u>	<u>%</u>	
States represented		28	56.0	
Configuration of district				
PreK-12		29	43.9	
K-12		35	53.0	
Other		2	3.0	
Type of district				
Urban		33	86.4	
Suburban		20	30.3	

program was implemented.

Rural	4	6.1
High-stakes testing district impact		
Implement high-stakes testing program	54	81.8
Test security breach within past two years	28	42.4

* Medians in addition to means are reported due, in part, to extreme outlier values in enrollments.

A simple bivariate correlation revealed no significant relationship to exist between district enrollment size and whether a test security breach occurred within the past two years ($r = .18; p > .05$). A significant correlation did emerge between whether the directors perceived that the district implemented a high-stakes testing program and whether a test security breach occurred ($r=.31; p < .05$). This indicated that school districts that tended to have high-stakes testing programs evidenced a tendency for test security breaches compared to those who did not have this type of program.

Prevalence of Identified Standards

Results indicate that test directors indicate that *The Standards* are most always prevalent within one's district (see Table 2). The highest rating ($\underline{M} = 4.7$) was given to standards associated with reasonable efforts made to assure the validity of test scores and making reasonable efforts to ensure that contrary, as well as supporting evidence is examined to settle the matter of the validity of the score in question. The overall lowest average rating ($\underline{M} = 4.3$) was for the standard indicating that should a test security irregularity be purported in my district, test-takers are given sufficient warning and an opportunity to provide evidence that scores should not be canceled or withheld. It is interesting to note that a few respondents wrote next to some standards

comments indicating that they did not know what the statement meant. This generally occurred next to statements associated with scoring mechanisms.

Table 2

**Frequency Distribution, Means, and Standard Deviations of
Test Directors' Perceived Level of Prevalence of Each Identified
Standard within the District**

		Always				Never		<u>M</u>	<u>SD</u>
		5	4	3	2	1			
1. Educators who administer and those responsible for high-stakes tests in my district take appropriate security precautions before, during, and after administration.	<u>n</u>	36	28	1	0	0	4.5	.53	
	%	54.5	42.4	1.5	0	0			
2. Educators in my district avoid any conditions in the conduct of the administration of high-stakes tests that might invalidate the results.	<u>n</u>	33	31	1	0	0	4.5	.53	
	%	50.0	47.0	1.5	0	0			
3. Reasonable efforts are made by educators in my district to assure the validity of test scores by eliminating opportunities for test-takers to attain scores by fraudulent means.	<u>n</u>	45	20	0	0	0	4.7	.47	
	%	69.2	30.8	0	0	0			
4. Reasonable efforts are made by students in my district to attain test scores by acceptable means.	<u>n</u>	31	30	3	0	0	4.4	.59	
	%	48.4	46.9	4.7	0	0			
5. Should scores in my district be withheld or canceled, the type of evidence and procedures used to determine this is fully explained to all test takers impacted.	<u>n</u>	38	10	8	3	0	4.4	.91	
	%	64.4	16.9	13.6	5.1	0			
6. Should a test security irregularity be purported in my district, test-takers are given advance warning and an opportunity to provide evidence that scores should not be canceled or withheld.	<u>n</u>	36	11	6	2	3	4.3	1.12	
	%	62.1	19.0	10.3	3.4	5.2			
7. Should a test security irregularity be purported in my district, reasonable efforts are made to ensure that contrary, as well as supporting evidence is examined to settle the matter of irregularity as well as validity of the score in question.	<u>n</u>	48	12	2	0	0	4.7	.51	
	%	77.4	19.4	3.2	0	0			

Importance of Actions, Procedures, and/or Policies

As noted within Table 3, test directors rated 22 of the 24 actions, procedures, and/or policies at least 4.0 on a five-point Likert scale used to measure importance (“5” meaning “extremely”). Those most important to deter test security breaches resulting in overall averages of 4.9 on a five-point scale were:

- All test materials are kept in locked, secured areas before and after test administration.
- All students eligible for testing are encouraged to participate in the testing program.
- Test directions related to time are strictly adhered to.
- Those deemed least important (though still having average ratings over 3.0) included:
 - There is a written policy with procedures that require the school superintendent to publicly report all breaches in test security (\underline{M} = 3.1).
 - Students take exams in only in designated classrooms (not in a large room such as a cafeteria, auditorium, or library) (\underline{M} = 3.8).
 - All proctors provide written verification that proper test administration procedures and security were adhered to (\underline{M} = 3.8).
 - Auditors routinely monitor test administration including adherence to test security (\underline{M} = 3.9).

Table 3

Frequency Distribution, Means, and Standard Deviations of Test Directors' Perceived Degree of Importance of the Existence of Actions, Procedures, and/or Policies to Deter Test Security Breaches

		Extremely				Not at all		<u>M</u>	<u>SD</u>
		5	4	3	2	1			
1. Students take exams only in designated classrooms (not in a large room such as a cafeteria, auditorium, or library).	<u>n</u> %	25 39.7	18 28.6	9 14.3	7 11.1	4 6.3	3.8	1.24	
2. All test proctors must be trained in test administration procedures, including test security.	<u>n</u> %	2 3.2	10 15.9	51 81.0	0 0	0 0	4.7	.47	
3. All proctors provide written verification that proper test administration procedures and security were adhered to.	<u>n</u> %	24 38.1	11 17.5	20 31.7	6 9.5	2 3.2	3.8	1.15	
4. All students are informed of obligations to abide by guidelines such as no unauthorized giving, receiving, or reproducing unauthorized materials or information.	<u>n</u> %	34 54.8	21 33.9	6 9.7	1 1.6	0 0	4.4	.74	
5. All test proctors are informed of acceptable responses to student questions posed during test administration.	<u>n</u> %	43 68.3	12 19.0	6 9.5	2 3.2	0 0	4.5	.80	
6. All test materials are kept in locked, secured areas before and after test administration.	<u>n</u> %	56 88.9	6 9.5	1 1.6	0 0	0 0	4.9	.47	
7. Supplemental materials are provided only during appropriately specified sections and are collected when completed.	<u>n</u> %	49 79.0	10 16.1	3 4.8	0 0	0 0	4.7	.54	
8. All students eligible for testing are encouraged to participate in the testing program.	<u>n</u> %	55 88.7	6 9.7	1 1.6	0 0	0 0	4.9	.38	
9. No ESL/LEP students are intentionally excluded from testing due to the potential impact of their test scores on accountability structures (such as overall school results).	<u>n</u> %	55 87.3	6 9.5	2 3.2	0 0	0 0	4.8	.45	
10. No special education students are intentionally excluded from testing due to the potential impact of their test scores on accountability structures (such as overall school results).	<u>n</u> %	54 85.7	6 9.5	3 4.8	0 0	0 0	4.8	.50	
11. All policies/procedures for regular testing are strictly adhered to during make-up testing sessions.	<u>n</u> %	54 87.1	7 11.3	0 0	0 0	1 1.6	4.8	.59	

Test Security In a High Stakes Environment: Myth Or Reality?

12. Test directions related to time are strictly adhered to.	<u>n</u> %	56 84.8	5 7.6	0 0	0 0	0 0	4.9	.28
13. There are specific written procedures that require proctors to report all breaches in test security.	<u>n</u> %	43 65.2	15 22.7	4 6.1	0 0	1 1.5	4.6	.76
14. There are written procedures that require building administrators to report all breaches in test security.	<u>n</u> %	47 71.2	11 16.7	3 4.5	2 3.0	0 0	4.6	.72
15. There are written procedures that require the district test director to report all breaches in test security.	<u>n</u> %	51 82.3	6 9.7	4 6.5	0 0	1 1.6	4.7	.73
16. There is a written policy with procedures that require the school superintendent to publicly report all breaches in test security.	<u>n</u> %	13 21.3	11 18.0	22 36.1	4 6.6	11 18.0	3.1	1.34
17. All educators involved with test materials are required to sign a security agreement	<u>n</u> %	32 51.6	13 21.0	8 12.9	4 6.5	5 8.1	4.0	1.29
18. There is a policy that specifically states repercussions of test security breaches by a student.	<u>n</u> %	26 42.6	18 29.5	14 23.0	1 1.6	2 3.3	4.1	1.01
19. There is a policy that specifically states repercussions of test security breaches by an educator.	<u>n</u> %	42 65.6	17 26.6	5 7.8	0 0	0 0	4.6	.64
20. There is a process whereby students and/or educators may confidentially report breaches of test security.	<u>n</u> %	30 48.4	22 35.5	9 14.5	0 0	1 1.6	4.3	.84
21. Auditors routinely monitor test administration including adherence to test security.	<u>n</u> %	26 40.6	14 21.9	18 28.1	4 6.3	2 3.1	3.9	1.10
22. The State Department of Education must provide clear written policies and procedures and guidelines regarding test security.	<u>n</u> %	56 87.5	5 7.8	3 4.7	0 0	0 0	4.8	.49
23. The State Department of Education must provide training of test administration, especially test security.	<u>n</u> %	44 68.8	11 17.2	6 9.4	2 3.1	1 1.6	4.5	.91
24. The State Department of Education must have a policy on repercussions for educators who breach test security.	<u>n</u> %	44 68.8	12 18.8	7 10.9	1 1.6	0 0	4.5	.75

Existence of Actions, Procedures, and/or Policies

As noted within Table 4, test directors rated the following actions, procedures, and/or policies most in existence within their districts to deter test security breaches:

- All students eligible for testing are encouraged to participate in the testing program ($\underline{M} = 4.6$).
- Supplemental materials are provided only during appropriately specified sections and are collected when completed ($\underline{M} = 4.5$).
- All test materials are kept in locked, secured areas before and after test administration ($\underline{M} = 4.5$).
- Test directions related to time are strictly adhered to ($\underline{M} = 4.5$).

Those deemed least important included:

- There is a written policy with procedures that require the school superintendent to publicly report all breaches in test security ($\underline{M} = 2.2$).
- Auditors routinely monitor test administration including adherence to test security ($\underline{M} = 3.0$).
- There is a policy that specifically states repercussions of test security breaches by a student ($\underline{M} = 3.1$).
- All proctors provide written verification that proper test administration procedures and security were adhered to ($\underline{M} = 3.2$).
- There is a process whereby students and/or educators may confidentially report breaches of test security ($\underline{M} = 3.3$).

Difference between Perceived Importance and Existence of Actions, Procedures, and/or Policies

Table 5 summarizes the actions, procedures, and/or policies that are perceived by test directors to have the biggest difference between perceived level of importance and existence. A statistically significant difference was found to exist between the perceived importance and existence ratings of 23 of the 24 statements. Most of these were significant at the $p < .001$. These results indicate a generally higher level of perceived importance for actions, guidelines and/or policies than perhaps that which has been in place. The only non-significant finding related to students taking exams

only in designated classrooms and not a large room such as a cafeteria, auditorium, or library.

Table 4

Frequency Distribution, Means, and Standard Deviations of Test Directors' Perceived Degree of Existence of Actions, Procedures, and/or Policies to Deter Test Security Breaches

		Extremely			Not at all		<u>M</u>	<u>SD</u>
		5	4	3	2	1		
1. Students take exams only in designated classrooms (not in a large room such as a cafeteria, auditorium, or library).	<u>n</u> %	20 31.7	26 41.3	12 19.0	2 3.2	3 4.8	3.9	1.03
2. All test proctors must be trained in test administration procedures, including test security.	<u>n</u> %	28 44.4	23 36.5	11 17.5	1 1.6	0 0	4.2	.80
3. All proctors provide written verification that proper test administration procedures and security were adhered to.	<u>n</u> %	23 36.5	6 9.5	9 14.3	10 15.9	15 23.8	3.2	1.63
4. All students are informed of obligations to abide by guidelines such as no unauthorized giving, receiving, or reproducing unauthorized materials or information.	<u>n</u> %	21 33.3	18 28.6	16 25.4	5 7.9	3 4.8	3.8	1.14
5. All test proctors are informed of acceptable responses to student questions posed during test administration.	<u>n</u> %	26 41.3	21 33.3	12 19.0	3 4.8	1 1.6	4.1	.97
6. All test materials are kept in locked, secured areas before and after test administration.	<u>n</u> %	38 60.3	20 31.7	5 7.9	0 0	0 0	4.5	.64
7. Supplemental materials are provided only during appropriately specified sections and are collected when completed.	<u>n</u> %	35 56.5	21 33.9	5 8.1	1 1.6	0 0	4.5	7.2
8. All students eligible for testing are encouraged to participate in the testing program.	<u>n</u> %	42 67.7	16 25.8	4 6.5	0 0	0 0	4.6	.61
9. No ESL/LEP students are intentionally excluded from testing due to the potential impact of their test scores on accountability structures (such as overall school results).	<u>n</u> %	36 57.1	19 30.2	7 11.1	1 1.6	0 0	4.4	.76
10. No special education students are intentionally excluded from testing due to the potential impact of their test scores on accountability structures (such as overall school results).	<u>n</u> %	36 57.1	13 20.6	12 19.0	1 1.6	1 1.6	4.3	.94

National Association of Test Directors 2001 Annual Proceedings

11. All policies/procedures for regular testing are strictly adhered to during make-up testing sessions.	<u>n</u> %	28 45.2	24 38.7	7 11.3	2 3.2	1 1.6	4.2	.89
12. Test directions related to time are strictly adhered to.	<u>n</u> %	39 63.9	19 31.1	3 4.9	0 0	0 0	4.6	.59
13. There are specific written procedures that require proctors to report all breaches in test security.	<u>n</u> %	33 52.4	12 19.0	9 14.3	4 6.3	5 7.9	4.0	1.29
14. There are written procedures that require building administrators to report all breaches in test security.	<u>n</u> %	39 61.9	6 9.5	8 12.7	5 7.9	5 7.9	4.1	1.34
15. There are written procedures that require the district test director to report all breaches in test security.	<u>n</u> %	42 67.7	5 8.1	7 11.3	3 4.8	5 8.1	4.2	1.30
16. There is a written policy with procedures that require the school superintendent to publicly report all breaches in test security.	<u>n</u> %	6 9.7	4 6.5	12 19.4	16 25.8	24 38.7	2.2	1.30
17. All educators involved with test materials are required to sign a security agreement	<u>n</u> %	28 45.2	6 9.7	8 12.9	7 11.3	13 21.0	3.5	1.60
18. There is a policy that specifically states repercussions of test security breaches by a student.	<u>n</u> %	17 27.9	8 13.1	12 19.7	12 19.7	12 19.7	3.1	1.50
19. There is a policy that specifically states repercussions of test security breaches by an educator.	<u>n</u> %	32 50.8	8 12.7	8 12.7	7 11.1	8 12.7	3.6	1.39
20. There is a process whereby students and/or educators may confidentially report breaches of test security.	<u>n</u> %	19 30.6	11 17.7	12 19.4	12 19.4	8 12.9	3.3	1.43
21. Auditors routinely monitor test administration including adherence to test security.	<u>n</u> %	14 21.9	11 17.2	12 18.8	12 18.8	15 23.4	3.0	1.48
22. The State Department of Education must provide clear written policies and procedures and guidelines regarding test security.	<u>n</u> %	34 53.1	12 18.8	12 18.8	2 3.1	3 4.7	3.7	1.29
23. The State Department of Education must provide training of test administration, especially test security.	<u>n</u> %	25 39.1	7 10.9	16 25.0	6 9.4	10 15.6	3.5	1.48
24. The State Department of Education must have a policy on repercussions for educators who breach test security.	<u>n</u> %	23 36.5	12 19.0	14 22.2	7 11.1	7 11.1	3.6	1.38

Table 5

Difference between Perceived Importance and Degree of Existence of Actions, Procedures, and/or Policies to Deter Test Security Breaches

	Impo r- tance	Exist - ence	t	Sign.
1. Students take exams only in designated classrooms (not in a large room such as a cafeteria, auditorium, or library).	3.8	3.9	-.49	ns
2. All test proctors must be trained in test administration procedures, including test security.	4.7	4.2	6.19	***
3. All proctors provide written verification that proper test administration procedures and security were adhered to.	3.8	3.2	4.61	***
4. All students are informed of obligations to abide by guidelines such as no unauthorized giving, receiving, or reproducing unauthorized materials or information.	4.4	3.8	5.34	***
5. All test proctors are informed of acceptable responses to student questions posed during test administration.	4.5	4.1	5.11	***
6. All test materials are kept in locked, secured areas before and after test administration.	4.9	4.5	3.68	***
7. Supplemental materials are provided only during appropriately specified sections and are collected when completed.	4.7	4.5	3.22	**
8. All students eligible for testing are encouraged to participate in the testing program.	4.9	4.6	3.39	**
9. No ESL/LEP students are intentionally excluded from testing due to the potential impact of their test scores on accountability structures (such as overall school results).	4.8	4.4	4.12	***
10. No special education students are intentionally excluded from testing due to the potential impact of their test scores on accountability structures (such as overall school results).	4.8	4.3	4.59	***
11. All policies/procedures for regular testing are strictly adhered to during make-up testing sessions.	4.8	4.2	6.40	***
12. Test directions related to time are strictly adhered to.	4.9	4.6	4.75	***
13. There are specific written procedures that require proctors to report all breaches in test security.	4.6	4.0	4.50	***
14. There are written procedures that require building administrators to report all breaches in test security.	4.6	4.1	4.44	***

15. There are written procedures that require the district test director to report all breaches in test security.	4.7	4.2	3.99	***
16. There is a written policy with procedures that require the school superintendent to publicly report all breaches in test security.	3.1	2.2	7.05	***
17. All educators involved with test materials are required to sign a security agreement	4.0	3.5	3.93	***
18. There is a policy that specifically states repercussions of test security breaches by a student.	4.1	3.1	6.10	***
19. There is a policy that specifically states repercussions of test security breaches by an educator.	4.6	3.6	4.58	***
20. There is a process whereby students and/or educators may confidentially report breaches of test security.	4.3	3.3	5.95	***
21. Auditors routinely monitor test administration including adherence to test security.	3.9	3.0	6.28	***
22. The State Department of Education must provide clear written policies and procedures and guidelines regarding test security.	4.8	3.7	7.14	***
23. The State Department of Education must provide training of test administration, especially test security.	4.5	3.5	6.50	***
24. The State Department of Education must have a policy on repercussions for educators who breach test security.	4.5	3.6	5.58	***

* $p < .05$. ** $p < .01$. *** $p < .001$, two-tailed.

Perceived Influence of High-stakes Testing

Over ninety percent of the test directors feel that over the past five years higher stakes have been attached to test results in their district. Most (*i.e.*, 88.6%) feel that higher stakes will be attached to test results over the next five years. With a five-point Likert-type scale and “5” indicating “strongly agree” and 1 “strongly disagree,” individuals gave moderate ratings to high-stakes testing enhancing student performance outcomes ($\bar{M} = 3.0$) and school improvement efforts ($\bar{M} = 3.4$). They also agree that “high-stakes testing does increase the potential of breaches in test security (see Table 6).

Table 6

Frequency Distribution, Means, and Standard Deviations of Test Directors’ Perceptions of the Impact of High-stakes Tests

		Extremely				Not at all		<u>M</u>	<u>SD</u>
		5	4	3	2	1			
1. Over the past 5 years, more “high-stakes” have been attached to test results in my district.	<u>n</u> %	51 79.7	9 14.1	2 3.1	2 3.1	0 0	4.7	.68	
2. In the next five years, more “high-stakes” will be attached to test results in my school district.	<u>n</u> %	44 66.7	14 21.9	6 9.4	0 0	0 0	4.6	.66	
3. “High-stakes” testing enhances student performance outcomes.	<u>n</u> %	5 7.8	19 29.7	21 32.8	11 17.2	8 12.5	3.0	1.14	
4. “High-stakes” testing enhances school improvement efforts.	<u>n</u> %	6 9.4	32 50.0	14 21.9	7 10.9	5 7.8	3.4	1.10	
5. “High-stakes” testing increases the potential of breaches in test security.	<u>n</u> %	38 60.3	20 31.7	3 4.8	2 3.2	0 0	3.5	.76	
6. Test security in a “high-stakes” environment can be a reality.	<u>n</u> %	27 46.6	25 43.1	4 6.9	2 3.4	0 0	4.3	.86	

Types of Test Security Breaches Experienced

A total of 25 (37.9%) individuals opted to provide a student-related incident and 34 (51.5%) an educator-related incident leading to a breach in test security. While each of these comments are provided in Tables 7 and 8, respectively, a thematic analysis yielded some interesting patterns. Student-related offenses tended to be tied to using of inappropriate materials, stealing secure materials, sharing answers, working on unallowable section of the test, and providing false identities. Educator-related offenses similarly included stealing and/or recording information from test secure materials. They also included coaching inappropriately, sharing and/or allowing students access to obtaining answers, deviating from published administration directions, and changing or assistance in modifying student answers.

Table 7

Examples Cited by District Test Directors of Educator-related Breaches in Test Security

- Student used the math reference sheet designed for the norm-referenced section of the test. The NRT portion of the test was invalidated.
 - A student from another district registered to take the test in the summer. The student had ID and registered as a male. The next year a student from the same school with the same ID registered as a female under the same name. The investigation is pending at the state level.
 - Several students were caught cheating on a mandated “high-stakes” test. Their scores were pulled from the total aggregate of scores at the site. The student’s parents were given a conference and the students themselves were disciplined with Saturday school.
 - Student tore out a page of a state math test and left the room before proctor realized what had happened. Student was located and page was returned. Student’s test was voided.
 - Evidence that students were passing on information about the content of a graduation demonstration exam (math) as evidenced by increase in mean scores by period (of day) during which exam was taken. Consequently we narrowed this testing window to a.m. only as opposed to all day.
 - Missing exam booklets – students never caught.
 - The student’s score was invalidated and throughout the district non-colored paper is provided for all students for scrap paper so that it is obvious if the student has brought in a sheet of paper or simply written down the information after coming to the testing session.
 - Lost writing prompt.
 - I subscribe to the Squirrel Principle; namely, that there is little we humans can do to keep a determined squirrel from reaching the bird feeder. I apply this to the issue of integrity of testing materials. Anyone who is determined to know the items on the test before administering the test can surely do that, no matter how securely we try and keep the booklets.
 - Cheating, if it occurs, would be handled by the school and result in the student being suspended.
 - Student test results have been invalidated if the student is discovered cheating or if the student does not complete the test.
 - These are handled at the school. My only quasi-involvement is when they steal a test.
 - Student worked on previously completed section of test. Test was invalidated. One student shared info with another. Even though it was not solicited, the test for the 2nd student was invalidated. Both situations involved the state “exit” exam.
 - Photocopied paper from a state assessment were used to instruct/prepare students. The copied material (5-8 pages) appears to be from a 1997 version of the test and their original has proved untraceable.
-

Table 7 (Continued)

Examples Cited by District Test Directors of Educator-related Breaches in Test Security

-
- A student left a testing room without permission and threw away his book. The student was punished. The test administrator and campus coordinator were reprimanded and new procedures were established.
 - Students were passing notes while a substitute teacher was proctoring the exam. The substitute read the note but did not realize that the material on the test was being discussed by the students. When the regular teacher returned, she realized what was going on and invalidated the student's exam.
 - High school teacher allowed students to take writing assessment material home. Principal notified test director who invalidated the results and notified state department. No consequences to teacher except maybe a verbal comment by principals.
 - Some students do not follow directions not to look forward or back in multi-part test. Students talked on the inventory testing.
 - Some test booklets were originally misplaced at a middle school but later they were located and returned to us. The consequences were fuzzy but the school had much frustration until the books were located.
 - Students were allowed to use calculators. A student reported. Teacher was questioned. Information found to be true; teacher reprimanded. Test scores pulled from classroom averages and site averages.
 - It was reported that two students shared answers when sitting next to each other. The proctor (there is both an examiner and proctor in each classroom) reported the incident to test coordinator who reported it to me. I talked to the principal. Children and proctor/examiner were interviewed. Children were administered an alternate form of sub-test affected. This allegation was referred to the central office for investigation. The office will determine if further investigation is necessary and if any disciplinary action is warranted.
 - Students were not monitored during a break in testing students and discussed test items.
 - A student grabbed a test and ran out of the classroom and building. A police officer went to the boy's house and confiscated the test. The boy was suspended.
 - State writing test has only one prompt. Students shared prompt with students testing during make-up period. These latter students were caught with draft essay. Scores were cancelled.
 - Student had contest-related material under desk during high school high-stakes test. Proctor "unaware" but observed by state inspector. Student results invalidated. Proctor reprimanded but without official sanction.
-

Table 8

Examples Cited by District Test Directors of Educator-related Breaches in Test Security

- Coaching and extended time were reported in one classroom. The principal confronted the teacher with the allegations. Although the teacher's accounting of events seem plausible, scores for that class will be closely examined when they arrive in the district.
 - The loss of a testing manual.
 - Teacher allowed use of student note material on test. Teacher resigned. Student scores allowed and materials would not have significantly aided them in testing.
 - A staff developer, teaching a writing strategy, asked the teachers in the group what the writing prompt was on the test just administered. The teachers told her. The prompt is embargoed for a certain time period and is not to be discussed until a certain date. The consultant was not aware of the rules for the writing test security.
 - An oral administration manual appeared to have been lost at a school. Materials were kept in locked storage. A thorough search was conducted. The manual was never found. Given the quality of materials that year it may never have been received. Materials checked in number ranges.
 - A principal was found to have utilized non-approved "prep" materials. The principal was reprimanded and a note placed in their personnel file.
 - Teacher copied questions out of a test prior to administration. Answer key was made and student answer documents were unofficially scored. Teacher was called to State Department for hearing. Credentials were suspended for one year.
 - Items on a vocabulary test were specifically taught to students in an elementary school. The incident was reported by principal. The student informed the principal that the words had been taught. The student produced class notes with word lists. Teacher was subsequently disciplined.
 - Writing prompt given early in year was "similar" to test. Teacher got prompt at a statewide workshop. Remedy: Test Coordinator warned not to use material unless they have documentation of where it was developed.
 - A test administrator was found to have been reading a student test booklet during the administration of the test. Statements were gathered from this individual and other adults in the testing room. After the investigation a formal report was made to the SDE and ultimately a letter of reprimand was sent to the teacher and placed in the personnel folder as well a warning that future violations of test security would cause teacher certification be revoked.
 - Teacher cheated though not fired. Reprimanded and transferred. She resigned.
 - A teacher released some tests that are considered "secure" to the district next door. This allowed teachers in the neighboring district to see/keep copies of tests they were not supposed to have.
-

Table 8 (Continued)

Examples Cited by District Test Directors of Educator-related Breaches in Test Security

-
- A teacher published secure high school course exams in a newspaper operated by a faction of the teachers' union. He was fired and the district sued him for copyright infringement. He counter-sued citing First Amendment rights. The counter-suit was thrown out by the judge and the infringement suit is still pending.
 - An educator has been reported to the state department of education. After the student turns in the test booklet, the teacher notices no responses in some areas. The test booklet was returned to the student with an encouragement to continue. The teacher looked ahead/read the math test and prepared the class by working with the class on those specific math problems the day before taking the math tests.
 - Teacher marked the test answer sheet of students who finished early and said "check" these you got them wrong! She did it repeatedly for some students until time was up. She lost bonus \$ and career ladder \$, about \$3-4K and was reprimanded. She was also transferred to another school.
 - Educator assisted students with written composition on a state test. Educator was suspended 5 days without pay by local action and received permanent reprimand on state teacher credential.
 - Building coordinator threw away damaged test booklets instead of returning them to district coordinator. Building coordinator's apparent inability to read and follow instructions has earned a letter of reprimand to the person's permanent file.
 - The coordinator of a school lost a set of field test booklets. After an extended search it was presumed the box they were in was picked up by a custodian. The box should have been in locked storage. Memos to the permanent files for the campus coordinators were written by the district's investigator and the incident was logged in the principal's evaluation.
 - Asking student to reconsider answers that are incorrect. Student retained at end of school day and asked to redo some work on tests. Insufficient evidence. Discussed allegation with teacher, principal, area supervisor, and test director.
 - One teacher took a copy of the test aside and used it to prepare her studies.
 - Teacher placed parallel test form items on overhead for student drill and discussion.
 - Xeroxing a student's constructed response so that when results are returned the teacher could see exactly what the student wrote (compare to the score received).
 - Tests were lost. Teachers review/examine test items during testing.
 - Some teachers were dismissed three years ago due to "test security" breaches. This was before I came to the district position of assessment.
 - Teacher did not follow time limits for sub-tests. Investigated teacher received reprimand. All student scores pulled from aggregated data. Parents were informed that data results did not reflect standardized administration.
-

Table 8 (Continued)

Examples Cited by District Test Directors of Educator-related Breaches in Test Security

- A teacher was reported to have provided correct answers to several children. The students were individually interviewed by the principal as to the incident. I interviewed the teacher. It was determined that the children should retake an alternate form of the test.
 - Item response data were analyzed by the central board. The findings have been presented at a disciplinary hearing in which it will be determined whether the teacher should be terminated.
 - Teacher left test booklet in classroom after testing. Booklet was lost.
 - Principal thought to have given students coaching/encouragement to “go on; you can do more”. Teacher told this was inappropriate and I collected tests. Teacher was disciplined after investigation.
 - Our state uses an NRT (commercial shelf product) originally implemented in 3 grades only. Publishing company helped passed out off level tests with no warnings. Over next two years more grades were tested and with higher stakes. Teacher used what she thought were sample items that turned out to be new. Teacher will probably be reprimanded for copyright violation by district. District will be cited for failure to use approved practice materials. The state and company deny workshops were ever held.
 - A new principal discovered that the test scores for a 3rd grade teacher from the previous year were unrealistically high (almost all most perfect scores). A joint investigation was conducted by the assessment office and principal that led to findings of fact. They were reported to employer and disciplinary process.
 - Changing marking and answer sheets. Reading a reading test to students.
 - Last spring a middle school math teacher used a homework/worksheet just before standardized testing. This compromised scores for about 130 students. We were able to work with our state department.
 - Alleged assistance in NRT administration – working, erasures, etc.
-

SUMMARY, DISCUSSIONS, AND IMPLICATIONS OF RESULTS

This study had sought to expand upon the previous work of Mehrens (1993) and NCME (1991) and others to examine district test directors’ perceptions of test security in a high-stakes environment. With surveys responses received from 66 individuals representing school districts from more than half the states and over 4.4 million students, results indicated that most are implementing high-stakes test program and

nearly half have experienced a breach in security the past two years. These directors indicate that practices recommended in *The Standards* (AERA *et al.*, 1985, 1999) are generally widely prevalent within their districts. The provided listing of 24 statements representing actions, guidelines, and/or policies that may be implemented to deter security breaches were deemed to be important by the test directors. Yet, level of prevalence did not meet the degree of established significance in 23 of these areas. There is strong agreement that the stakes placed on testing has risen the past five years and they project for this trend to continue into the next five years. Thematic analyses of open-ended comments indicated that both student and educator-related offenses included stealing and/or recording information from secure materials. Students also tended to use inappropriate materials, share answers, work on unallowable sections of the test, and present false identities. Educator-related offenses included coaching inappropriately, sharing and/or allowing students access to obtaining answers, deviating from published administration directions, and changing or assistance in modifying student answers. Not surprisingly, there was found to be a statistically significant relationship between those implementing high-stakes testing programs and breaches in security.

The results of this study may be pertinent to test developers, test directors, test administrators, and test users. By gaining a better understanding of what test directors feel are important steps to maintaining test security, individuals can seek to provide and/or implement those steps which have been identified. Additionally, one may relate the emerging themes of types of student and educator breaches to the associated

actions, policies, and/or guidelines in a proactive method to help deter cheating.

Implications are also realized for professional development training. For example, one may use the vignettes presented within this report as part of discussion groups held on enhancing test secure practices. Researchers may wish to further explore the types of actions taken by districts deemed most successful in deterring these types of breaches. Additionally, with the perception that high-stakes testing will continue to increase in the coming years, on-going research should continue to track patterns over time and to develop and assess new methods to continue to deter breaches such as the use of technology.

Appendix A: Cover Letter Accompanying Survey



Dear NATD Colleague:

With emphasis being placed on high-stakes accountability testing throughout the United States, much attention has been given to test security. But, is test security in a high-stakes environment a myth or can it be a reality? What are the perceptions of those of us who work as testing directors in school districts?

This topic of test security is of interest to those of us who are responsible for facilitating testing programs in today's high-stakes testing environment. As such, the purpose of the enclosed survey is to collect information and investigate the perceptions of test directors about test security, particularly in a high-stakes testing environment. Results may then be used to help improve and enhance test security practices in the future.

As a testing director, **your perspective is essential** and will help develop solutions in the area of high-stakes testing. Thus, your participation in completing and returning the attached survey is most important. The results of this study will be incorporated as one component of our organization's invited symposium during the National Council Measurement in Education conference in Seattle, Washington, in April and will be posted on the NATD homepage at <http://www.natd.org>.

I would appreciate it if you would return your completed survey by **MARCH 18** via FAX to my attention at (757) 638-3303.

Thank you in advance for completing this important survey in a timely manner and I look forward to sharing the results with you.

Sincerely,

Mary E. Yakimowski

NATD Vice President

PART III: TEST SECURITY PRACTICES

For each of the statements below, please answer two questions. First, how important is this practice to you in maintaining test security? Second, how much do you feel this practice routinely exists within your district? In rating the importance, use the five-point scale with “5” indicating “extremely important” and “1” indicating “not important at all.” To indicate degree practice exists in your school district, use the five-point scale with “5” indicating “strongly agree” and “1” indicating “strongly disagree.”

	<u>How important?</u>					<u>Agree it is existent?</u>				
	<i>Extremely Important</i>				<i>Not important at all</i>	<i>Strongly Agree</i>				<i>Strongly Disagree</i>
1. Students take exams in only in designated classrooms (not in a large room such as a cafeteria, auditorium, or library).	5	4	3	2	1	5	4	3	2	1
2. All test proctors must be trained in test administration procedures, including test security.	5	4	3	2	1	5	4	3	2	1
3. All proctors provide written verification that proper test administration procedures and security were adhered to.	5	4	3	2	1	5	4	3	2	1
4. All students are informed of obligations to abide by guidelines such as no unauthorized giving, receiving, or reproducing unauthorized materials or information.	5	4	3	2	1	5	4	3	2	1
5. All test proctors are informed of acceptable responses to student questions posed during test administration.	5	4	3	2	1	5	4	3	2	1
6. All test materials are kept in locked, secured areas before and after test administration.	5	4	3	2	1	5	4	3	2	1
7. Supplemental materials are provided only during appropriately specified sections and are collected when completed.	5	4	3	2	1	5	4	3	2	1
8. All students eligible for testing are encouraged to participate in the testing program.	5	4	3	2	1	5	4	3	2	1
9. No ESL/LEP students are intentionally excluded from testing due to the potential impact of their test scores on accountability structures (such as overall school results).	5	4	3	2	1	5	4	3	2	1
10. No special education students are intentionally excluded from testing due to the potential impact of their test scores on accountability structures (such as overall school results).	5	4	3	2	1	5	4	3	2	1
11. All policies/procedures for regular testing are strictly adhered to during make-up testing sessions.	5	4	3	2	1	5	4	3	2	1
12. Test directions related to time are strictly adhered to.	5	4	3	2	1	5	4	3	2	1
13. There are specific written procedures that require proctors to report all breaches in test security.	5	4	3	2	1	5	4	3	2	1
14. There are written procedures that require building administrators to report all breaches in test security.	5	4	3	2	1	5	4	3	2	1
15. There are written procedures that require the district test director to report all breaches in test security.	5	4	3	2	1	5	4	3	2	1
16. There is a written policy with procedures that require the school superintendent to publicly report all breaches in test security.	5	4	3	2	1	5	4	3	2	1

	<u>How important?</u>					<u>Agree it is existent?</u>				
	<i>Extremely Important</i>					<i>Not important at all</i>	<i>Strongly Agree</i>			<i>Strongly Disagree</i>
17. All educators involved with test materials are required to sign a security agreement	5	4	3	2	1	5	4	3	2	1
18. There is a policy that specifically states repercussions of test security breaches by a student.	5	4	3	2	1	5	4	3	2	1
19. There is a policy that specifically states repercussions of test security breaches by an educator.	5	4	3	2	1	5	4	3	2	1
20. There is a process whereby students and/or educators may confidentially report breaches of test security.	5	4	3	2	1	5	4	3	2	1
21. Auditors routinely monitor test administration including adherence to test security.	5	4	3	2	1	5	4	3	2	1
22. The State Department of Education must provide clear written policies and procedures and guidelines regarding test security.	5	4	3	2	1	5	4	3	2	1
23. The State Department of Education must provide training of test administration, especially test security.	5	4	3	2	1	5	4	3	2	1
24. The State Department of Education must have a policy on repercussions for educators who breach test security.	5	4	3	2	1	5	4	3	2	1

PART IV: PERCEPTIONS OF TEST SECURITY OVER TIME

Please use this five-point scale to circle your response to indicate the extent to which you agree with the statement.

STRONGLY AGREE	AGREE	NEUTRAL	DISAGREE	STRONGLY DISAGREE
5	4	3	2	1

1. Over the past 5 years, more “high-stakes” have been attached to test results in my district.	5	4	3	2	1
2. In the next five years, more “high-stakes” will be attached to test results in my school district.	5	4	3	2	1
3. “High-stakes” testing enhances student performance outcomes.	5	4	3	2	1
4. “High-stakes” testing enhances school improvement efforts.	5	4	3	2	1
5. “High-stakes” testing increases the potential of breaches in test security.	5	4	3	2	1
6. Test security in a “high-stakes” environment can be a reality.	5	4	3	2	1

PART V: TEST SECURITY BREACHES *(Please feel free to use the reverse side to complete these questions.)*

1. Please briefly describe a student-related test security breach that has occurred in your district. Share some of the specifics related to the breach along with the consequences. *(Do not use the student’s name. Also, please note that anonymity of your name and that of the district’s will be assured.)*
2. Please briefly describe an educator-related test security breach that has occurred in your district. Share some of the specifics related to the breach and the consequences. *(Please do not use the educator’s name. Also, please note that anonymity of your name and that of the district’s will be assured.)*

REFERENCES

- American Educational Research Association. (1985). *AERA position statement concerning high-stakes testing in preK-12 education*. Available: <http://www.aera.net/about/policy/stakes.htm>
- American Educational Research Association, American Psychological Association and National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Educational Research Association, American Psychological Association and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. (2nd ed.). Washington, DC: American Psychological Association.
- Andrews, C. (2001, March). Stanford test: Questions never change. *Alabama Live*. Available: <http://www.al.com/news/mobile/Jul2000/27-a390735a.html>
- Borja, R. R. (1999, April). Henrico officials find security breach in SOLs. *Times Dispatch*. Available: <http://www.discoverrichmond.com/rtd/special/sol/sol08.shtml>
- Bradley, A. (1999, February). Chicago Schools take aim at teacher newspaper. *Education Week on the Web*. Available: <http://www.edweek.org/ew/vol-18/22chic.h18>
- Bushweller, K. (1999, April). Who's cheating in our schools? Our survey results might alarm you. *American School Board Journal*. Available: <http://www.asbj.com/199904/0499inprint.html>
- Bushweller, Kevin. (1997, September). Teaching to the test. *The American School Board Journal*. Available: <http://www.asbj.com/1997/asbj0997.html>
- Canner, J. (1992). Regaining the public trust: A review of school test programs, practices. *NASSP Bulletin*, 76(545), pp. 6-15.
- Cizek, G. J. (1999) *Cheating on tests: How to do it, detect it, and prevent it*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Colton, G. D. (1997). *High-tech approaches to breeching [sic] examination security: Espionage 101*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.
- Corbett, H. D. & Wilson, B. (1991). *Testing, reform, and rebellion*. Norwood, NJ: Ablex.

Danitz, T. (2001, February). States pay \$400 million for tests in 2001. *Stateline.org*. Available: <http://www.stateline.org/story.cfm?storyid=16627>

Elmore, R. F., Abelman, C. H., & Fuhrman, S. H. (1996). The new accountability in state education reform: From process to performance. In H. F. Ladd (Ed.). *Holding schools accountable: Performance-based reform in education* (pp. 65-98). Washington, DC: The Brookings Institution.

Gordon, B. M. (2000, Fall). On high-stakes testing. *American Educational Research Association Division G Newsletter*. Available: <http://www.aera.net> front page

Grant, S. G. (2000). Exploring teachers' perceptions of changes in the New York State testing program. *Education Policy Analysis Archives*. [On-line serial]. 8(4). Available: <http://epaa.asu.edu/epaa/v8n14.html>

Hall, J. L. & Kleine, P. F. (1992). Educators' perceptions of NRT misuse. *Educational Measurement: Issues and Practice*, 11(2) 18-22.

Haney, W., Madaus, G. M., & Lyons, R. (1993). *The fractured marketplace for standardized testing*. Boston: Kluwer Academic Publications.

Hoff, D. J. (2001, March). States spend nearly half a billion on testing. *Education Week*.

Hopkins, A. (1999, March). AIMS field test suffers breach of security. *Foothills News*. Available: <http://www.ahwatukee.com/afn/weekend/breaking/articles/990319c.html>

Kerlinger, F. N. (1986). *Foundations of behavioral research*. NY: Holt, Rinehard and Winston, Inc.

Kher-Durlabhji, N. & Lacina-Gifford, L. J. (1992, April). *Quest for test success: Preservice teachers' vies of "high-stakes" tests*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Knoxville, TN.

Kilian, L. J. (1992). A school district perspective on appropriate test-preparation practices: A reaction to Popham's proposals." *Educational Measurement: Issues and Practices*, 12-15.

Kimmel, E. W. (1997). *Unintended consequences or testing the integrity of teachers and students*. Paper presented at the Annual Assessment Conference of the Council of Chief State School Officers, Colorado Springs, CO.

Ligon, G. (1985). *Opportunity knocked out: Reducing cheating by teachers on student tests*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 263 181).

Ligon, G. (2000). *Test tactics: Trouble with a capital "T"*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Linn, Robert L. (2000a). Assessments and accountability. *Educational Researcher*, 29(2), 4-16.

Linn, Robert L. (2000b) Assessment and accountability (Condensed version). *Practical Assessment, Research and Evaluation*, 7(11). Available: <http://ericcae.net/pare/getvn.asp?v=7&n=11>.

Mehrens, W. A. (1993). Survey of test security practices. *Educational Measurement: Issues and Practices* 12(4) 5-19.

Mehrens, W. A. (1998). Consequences of assessment: What is the evidence? *Education Policy Analysis Archives*. [On-line serial]. 6(13). Available: <http://epaa.asu.edu/epaa/v6n13.html>

Mehrens, W. A. & Lehman, I. J. (1991). *Measurement and evaluation in education and psychology*. (4th ed.). Fort Worth, TX: Holt, Rinehardt and Winston.

Monsaas, J. A. & Engelhard, G., Jr. (1991). *Attitudes toward testing practices as cheating and teachers' testing practices*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

National Council on Measurement in Education. (1991, April). *Regaining trust: Enhancing the credibility of school testing programs: A report from an NCME task force*. Washington, DC: NCME.

O'Reilly, J. M. (Undated). *Raising standardized test scores*. Mesa, AZ: Assessment and Special Projects, Mesa School District.

Orlans, H. (2000, July). *Improving student performance*. Heldref Publications. Available: http://www.findarticles.com/cf_0/m1254/4_32/64189891/print.jhtml

Perlman, C. (2000). Surreptitious inclusion of good teaching in test preparation activities. *ERS Spectrum*. 12-20.

Phelps, R. P. (1999). Why testing experts hate testing. *Fordham Report*, 3(1). Available: <http://www.edexcellence.net/library/phelps.htm>

Phelps, R. P. (2000). The extent and character of system-wide student testing in the United States. *Educational Assessment*, 4(2), 89-121.

Picus, L. O. & Tralli, A. (1998). Alternative assessment programs: What are the true costs? *CSE Technical Report 441*, Los Angeles: CRESST, p. 47.

Popham, W. J. (2000a). The score boosting game. *American School Board Journal*, 187, 36-39.

Popham, W. J. (2000b). *Teaching to the test: High crime, misdemeanor, or just good instruction?* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Popham, W. J. (1999). Where large scale educational assessment is heading and why it shouldn't. *Educational Measurement*, 18(3).

Protheroe, N. (2000). Ethical and appropriate test preparation. *Educational Research Services The Informed Educator Series*.

Ravitch, D. (Ed.). (1995). *Debating the future of American education: Do we need national standards and assessments?* Washington, DC: Brookings Institution.

Resnick, D. & Resnick, L. (1985). Standards, curriculum, and performance: A historical and comparative perspective. *Educational Researcher*, 14(4), 5-20.

Sandham, J. L. (1999, April). Exam-testing breaches put focus on security. *Education Week on the Web*. Available: <http://www.edweek.org/ew/vol-18/30secure.h18>.

Schwartz, J. L. & Viator K. A. (Eds.) (1990). *The process of secrecy: The social, intellectual, and psychological costs of current assessment practice*. A report to the Ford Foundation.

Shepard, L. A. (2000). *The role of assessment in a learning culture*. Presidential address presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. Available: <http://www.aera.net/meeting/am2000/wrap/praddr01.htm>

Shepard, L. A. & Dougherty, K. C. (1991). *Effects of high-stakes testing on instruction*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 337 468).

Smith, M. L. (1991). Meanings of test preparation. *American Educational Research Journal*. 28(3), 521-42.

Webster, T. (2000, October). Cheating on high school tests alleged. *Las Vegas SUN*. Available:
<http://www.lasvegassun.com/sunbin/stories/archives/2000/oct/25/510948616.html>

Wellhousen, K. & Martin, N. K. (1995). Preservice teachers and standardized test administration: Their behavioral predictions regarding cheating. *Research in the Schools*. 2(2) 47-49.

Wiersma, W. & Jurs, S. G. (1990). *Educational measurement and testing* (2nd ed.). (1990). Allyn and Bacon, Needham Heights, MA.

Wilson, H. E. (1993). Ethics in education. *School Management Advisor*. N 21. (Eric Document Reproduction Service No. ED375155).

ⁱ. The developers of the indices g_2 and ω have made programs available for calculating those indices. Readers wishing to obtain a program for calculating g_2 should send a request via email to Robert Frary at fraryrb@prodigy.net; users interested in a program to calculate ω should send an email request to James Wollack at jwollack@facstaff.wisc.edu.

ⁱⁱ. The Stancik report itself has not been without controversy. An investigation and report on Stancik's findings commissioned by the teachers' union (United Federation of Teachers) and conducted by Thacher and Associates (2000) was highly critical of the methods employed in the initial investigation. The report concludes that the original investigation may have incorrectly identified some educators as having engaged in inappropriate practices.

Cheating on Large Scale Assessments: A Publisher's Perspective

Michael Kean

CTB/Mc-Graw-Hill

Consider the following quotations:

From the *Cleveland Plain Dealer*: "According to interviews with a dozen local teens, some test aids were taped onto the legs of school girls, just above their shirts' hemlines. Others were placed in folders with see-through covers, then plopped on desks in plain view. Many cheat sheets were attached to pens or the insides of water

bottles or slipped into calculator covers. All are techniques perfected during tests taken throughout the year.”

Or this, from the *Los Angeles Times*: “The State Department of Education has determined that adults apparently altered students’ standardized test results in two classes at a Pasadena elementary school to improve scores... A high number of answers in two, third-grade classes were changed from wrong to right, an analysis determined. The findings will invalidate the school’s gains.”

A final example: “...tomorrow is the first day of the examination...it consists of a long passage which the candidate has not seen. This passage is printed on the examination paper, and it would naturally be an immense advantage if the candidate could prepare for it in advance. For this reason, great care is taken to keep the paper secret.”

The report continues by describing how a student gained access to the room where the examination paper was kept, and copied it, the day before its formal administration.

Unlike the previous two quotations, in this instance we are not dealing with a state assessment program, but rather a competition for the prestigious Fortescue Scholarship. It was not a multiple-choice or a constructed response exam, but a performance assessment. The students were required to expound on a large passage of Greek translation, in this case from the historian Thucydides.

Unlike the two previous examples, this was not published earlier this year, but in June 1904 in London’s *Strand Magazine*. The source was “The Adventure of the Three Students,” a story by Sir Arthur Conan Doyle. Fortunately, in this instance, Sherlock Holmes happens to be in Cambridge at the time and solves the mystery. Unfortunately for us, the great detective has retired and is currently unable to help State Departments of Education and local education agencies deal with similar problems.

I use this example to make the simple point that cheating on high stakes tests is certainly not a new phenomenon.

The focus of this presentation will provide a publisher's perspective on cheating on large-scale assessments. In addition to providing a general perspective, I will also comment on the prevalence of the problem, what I believe to be the real issue, and what can be done to reduce the problem.

A disclaimer before I begin: I am providing you with **one** publisher's perspective, and though I suspect that many other publishers may come to similar conclusions, it is important to point out that I alone am responsible for my comments.

PERSPECTIVE

The topic of this symposium is "Cheating on Large-Scale Assessments." Large-scale assessments virtually always rely upon some type of standardized test. Not necessarily a norm referenced test, because the majority of state large-scale assessments are customized, though they may include a norm-referenced portion.

What is a standardized test? Simply a test developed, administered, and scored under standard conditions. Hence, we do not want to **d**estandardize any of the conditions related to the way we administer or score the test. If inappropriate activities occur, the test may become **d**estandardized and the information provided by it **d**evalued.

When can such activities occur?

1. Prior to the test being administered (as part of test preparation).
2. During the administration of the test.
3. After the administration of the test, both prior to and during the scoring process.

Allow me to provide some examples:

1. There have been numerous articles written concerning test preparation. Among the most useful are Mehrens' and Kaminski's 1989 piece in *Educational Measurement: Issues and Practices*. Many of you may be familiar with it and may be equally familiar with the seven-point continuum which runs from the most ethical (giving general instruction on direct objectives without referring to the objectives that the standardized test measure), to the most **un**ethical (providing practice or instruction on the test itself).
2. With respect to activities relating to the actual administration of a standardized test, virtually all test publishers provide very explicit guidelines which accompany the test materials. Similarly, customized state assessments are virtually always accompanied by even more specifically focused guidelines, and in many instances the tests are administered under semi-secure or secure conditions. Also included in both publishers' and states' guidelines are instructions relating to test accommodations for disabled and LEP students.
3. With respect to what happens after the administration of the test (both prior to and during the scoring process), it is important to note that there may be considerable variability between what a local school district using an off-the-shelf standardized achievement test does, and how that same district handles a mandated statewide assessment. The reason for these differences is that the nationally normed standardized test may be scored by the test publisher, by the local district itself, or sent to a third party for scoring. This opens numerous possibilities for "non-standard" activities.

With virtually all large-scale state assessments, however, all tests are scored uniformly, and very explicit instructions, including those relating to collecting, checking, packaging, and transporting the tests and/or answer sheets, are usually included.

So with respect to this publisher's perspective, the easiest way to deal with the situation is fairly simple:

1. If you think an activity may be inappropriate, don't do it.
2. Make sure you carefully read the guidelines before doing anything.
3. Carefully follow the guidelines.

HOW PREVALENT IS THE PROBLEM

I've already provided you with a number of recent and not so recent quotations.

Consider one more if you will:

- ◆ From the *Sacramento Bee*: "It was the hottest series of numbers to hit California since the \$104 million super lotto. Up and down the state this week, people clamored to see California's first academic ranking of its nearly 7,000 schools... But in the new high stakes educational climate where scholarships for kids, bonuses for teachers, and job security for principals are all on the line—there is a potential for something completely unintended: cheating. And it's not just for kids anymore. It happened in New York, it happened in Texas, and now it's happened in Southern California, where thirteen teachers were found this month to have shared exam questions with their students before last spring's standardized testing."

We have all seen the blazing headlines and the news articles about cheating. Bear in mind, however, that newspapers don't play up honesty. What you see or hear reported is not necessarily indicative of what is happening throughout the United States. What it does represent is those instances where cheating has occurred. What you don't hear reported are the 99 percent of test administrations that go as intended.

In some instances, the fear of cheating has reached such historic proportions that it's beginning to sound like the McCarthy era all over again. Consider what recently

occurred in New York City. A year after the City's special schools investigator issued a finding that dozens of teachers and two principals in thirty-two schools had helped students cheat on standardized tests, a former school inspector who reinvestigated the allegations determined that twenty-three of those teachers had, in fact, done nothing inappropriate.

Another personal example might also prove useful. As some of you know, I had ultimate responsibility for the School District of Philadelphia's assessment program through much of the 1970s. During those days we annually tested virtually every student in the School District. That's in kindergarten through grade twelve—approximately a quarter of a million students per year. During my eight and a half years heading the Office of Research and Evaluation, we discovered only two instances of deliberate attempts to cheat. One involved a principal who endeavored to put pressure on some of his teachers, and who was turned in by the teachers in that school. The other instance was a single teacher (in another school several years later) who provided the answers to her students. In this case, when it was discovered by the other teachers, they reported her to the central administration.

What is my conclusion concerning the prevalence of the problem? Given the tremendous increase of large-scale assessment, the problem is indeed more prevalent than a decade ago, but not nearly as common as the media would have us believe.

THE REAL ISSUE

Psychologists tell us that there are some individuals who are just naturally dishonest. The most publicized kleptomaniacs have been extremely wealthy individuals. These findings aside, however, it is extremely rare that individuals risk cheating on a test that doesn't have consequences. At the risk of having everyone in the audience say "duh," I'll state the obvious. "It's not the scale, it's the stakes." Large-scale testing is not the

culprit; it is large-scale testing with high stakes attached to the results, that, in some instances, has resulted in cheating.

I would further suggest to you that in these instances what has occurred is that the test results are being viewed as an end unto themselves rather than as information about achievement.

What Can Be Done To AVOID THE PROBLEM

Imagine, if you will, a solution spectrum. At one end, there are a number of solutions that might fall into the “get tough” category. These include (1) better security, (2) using technology to catch the perpetrators, and (3) stiffer penalties for those who are caught. Will these help? Probably, to some extent.

Of course, at the other end of the testing continuum there are groups such as “No Test,” which suggest fixing the problem by doing away with tests altogether. Since this is quite unlikely in the current political, social, and educational climate, we might wish to consider a middle ground. This middle ground calls for placing the role of assessment in context.

For example, Joe Nathan, Director the University of Minnesota’s Center for School Change based at the Humphrey Institute of Public Affairs, has found that there are nine key elements that schools, schools districts, and states can employ in designing accountability and assessment programs. Nathan indicates that six of these elements are **vital** in assuring appropriate approaches. They are:

- ◆ Clear and measurable goals for each school.
- ◆ Goals that are understood and supported by families, students, and educators.
- ◆ A variety of measures, including standardized tests and performance measures.

- ◆ The measurement of all students' work, including that of special education students.
- ◆ Assessments that measure the progress of students who don't speak English at home.
- ◆ The use of assessment information to shape school improvement efforts.

Among the other three elements, which Nathan deems “valuable” (but not vital), is one which I should suggest would be in the earlier category. And that is: “Creating a parent/educator/community committee to supervise assessment efforts.” I personally think that that such a group would be invaluable in helping forestall any irregularities in local assessment programs.

I would also like to suggest that we pay more attention to a series of recommendations made by the National Research Council report on assessment, and to the most recent revision of the APA/AERA/NCME *Standards for Educational and Psychological Testing*. In particular, states or local districts that make use of high stakes assessments should be sure that:

1. Students receive adequate notice of the test and its consequences.
2. Students have an opportunity to learn the knowledge and skills being tested, meaning that the test must be aligned with the curriculum.
3. The process for setting passing scores should be documented and evaluated.
4. Students should have equal access to any specific preparation for taking the test.
5. Students who risk failing the test should be advised in advance and provided with appropriate instructional remedial help that will improve their chances of passing it.
6. Students should have multiple opportunities to retake the test.

Finally, let me suggest an area that receives a great deal of lip service and little else. And that area is training, especially training teachers, principals, education policy makers, and the media to understand the appropriate role of assessment in education.

We could probably spend an entire session talking about the paucity of assessment training and the resulting misuse of assessment information. But that's a topic for yet another time, however.

CONCLUSION

In conclusion, let me state that CTB/McGraw-Hill, and I suspect most other test publishers, publish a list of explicit practices and procedures as part of their test administration guides. Such procedures reflect general ethical principles and standards, as typically reflected in a variety of well-known documents and position statements published by such organizations as AERA, APA, and NCME. Procedures reflect such areas as test security, pre-test activities (which include test preparation activities), testing conditions (including specific testing procedures, testing in the classroom, testing materials, directions, and monitoring), and post-test activities.

Test publishers and their partners—the state departments of education and local school districts with which they work—have collaborated closely in the past, and continue to do so today in the formulation of explicit procedures to guide the administration of all assessments, including large-scale assessments, especially those with high-stakes consequences.

Kentucky Test Security

**Linda Fraser, Robert E. Wetter
and Sharon Hutchinson**

Kentucky Department of Education

This paper was presented in PowerPoint format. The PowerPoint slides are presented on the following pages. The appendices and other materials discussed as part of the presentation follow the PowerPoint slides.

Overview

- I Test Construction Security
- II Contractor Security
- III Shipping and Receiving Security
- IV Testing Security
- V Scoring Security
- VI Writing Portfolio Ethics
- VII Allegations

I. Test Construction Security

- Teachers Construct All Test Items
- Non-disclosure Agreement Signed
- No Retained Materials
- Time Lag Before Items Used
- Not all Constructed Items Used

II Contractor Security

- Customary Contractor Internal Security
- Secure FTP Site for Data Transfer Between Contractors and SEA
- Every Test Booklet Bar-coded
- Non-Disclosure Agreements
- Secure SEA Server

III Shipping/Receiving Security

- Secure Storage of Testing Materials
- Test Booklet Counts at both Shipping and Receiving
- Shipped Boxes Recorded by School
- District Assessment Coordinator's Certification
- Principal's Certification
- SEA Inspection

IV Testing Security

- Professional Ethics
 - 704 KAR 20:680
- Educational Defensibility
 - Teach the content
- Student Ownership
 - All work done entirely by the student
 - Note inclusion of writing portfolio materials

IV Test Security -

Appropriate Assessment Practices

- Requires Training
- Read, Sign and Comply
- Must occur prior to portfolio development
- References "documents and administration manuals specific to state-required assessment components"

IV Test Security

- Central Office, School Office, Classroom
- "Shall not be reproduced in whole, in part or paraphrased in any way."
- Destroy notes, rough drafts created by students
- Ensure reusable materials are unmarked
- Don't share information about items

IV Test Security

- Use released items for Scrimmage Testing
- Security issues with students using technology
- Alert papers
- Procedures for Reporting Errors

IV Test Security

Classroom Materials

- Dictionaries/thesauri for On-demand Writing test ONLY
- Calculators for KCCT mathematics and one section of CTBS mathematics
- "Shall not distribute...any information or materials that are not sent...or specified..."

IV Test Security

Administration Practices

- Double check students with disabilities IEPs---accommodations? modifications?
- Lunch, interval, restroom breaks
- Same grade, same section, same time
- Illness during assessment session
- Supervise and monitor students

VI Writing Portfolio Development

Teachers may:



Ask questions to clarify, indicate position of errors, share holistic scoring guide



Teachers may not:

Direct corrections, revisions, working after completion date

IV Test Security

Disciplinary Practices and Student Motivation

- Re-administer for disciplinary practices
 - Original work is submitted for scoring
- Good Faith Effort
 - Checklists
 - No evaluative statements
 - Follow School Food Services Guidelines and the Finance "Redbook"

V Scoring Security

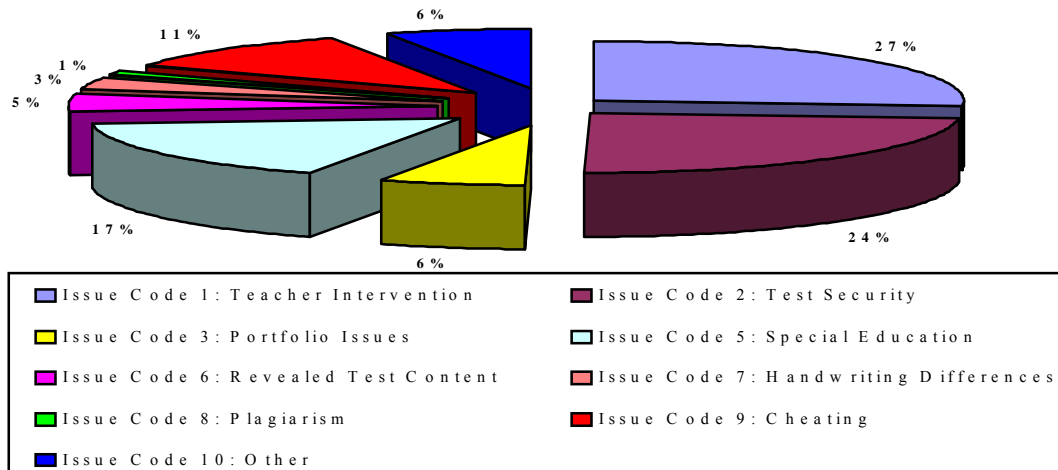
- Scoring by out of state, private contractor
- Scorers are informed of secure nature of assessment
- Scorers sign non-disclosure agreements
- Materials at scoring site are kept secure, and logged in and out of the scoring session
- No materials leave scoring site room
- SEA onsite monitoring

VI Writing Portfolio Scoring

- Certified staff
 - Teacher or administrator certification
 - Employed by district
 - Certified or classified position or leave of absence
 - Current year training

VII Allegations - 1999

Percentage of CATS/KIRIS Issues in 1998 & 1999



Typical Test Security Issues

- Test administered out of order/sequence
- Extended time on test/portfolios
- Students worked ahead
- Students allowed to mingle/go to lunch before finishing test
- Students allowed to re-enter portions of completed test
- Teacher gave unfinished test back to student to complete responses
- Student left school during testing period and allowed to complete test upon return
- Student took test materials out of testing area
- Test booklet lost/missing
- Teacher took notes from test booklet

Typical Teacher Intervention Issues

- Teacher allowed students to use dictionaries
- Teacher allowed students to use a "four-column method" organizational tool
- Teacher allowed calculators on CTBS tests
- Teacher answered test questions
- Teacher gave inappropriate assistance during testing
- Teacher marked/made corrections on response booklet
- Teacher coached answers

Typical Special Education Issues

- Student received accommodations not entitled to
- Student allowed to re-do a portion of the test
- Inappropriate prompting from scribe
- Teacher provided answers
- Teacher wrote responses for student
- Teacher/Scribe corrected words/spelling
- Teacher rewrote student responses
- Student allowed to type responses
- Student given extended time
- Teacher/Scribe allowed copying
- Teacher/Scribe gave inappropriate assistance
- Student without IEP given special accommodations

Only 0.2% Score Correction

Year	Allegations or Investigations	Substantiated Incidents	Number of Students with Changed Scores
1998	96	53	1275
1999	137	60	1220
2000	184	62	1081

Test Security

State Level Items from NATD Survey *Test Directors' Perceptions of Test Security*

- 1. The State Department of Education must provide clear written policies and procedures and guidelines regarding test security.**

WRITTEN POLICY

Kentucky has several things. The strongest is the Administration Code that has been incorporated as a state regulation. In addition to that regulation we have administration manuals, district assessment coordinator, principal, and test administrator certification forms. On those forms it stipulates what may and what may not be done. Those individuals have to sign off that those things were done appropriately.

AUDITORS' PROCEDURE

During the two week testing window teams of two visit schools chosen as random by the state for a monitoring visit. No warning is given. The state department person just pops in and asks questions. A specific **questionnaire is used at the district office w/ District Assessment Coordinator (DAC) and Superintendent of Instruction. What happens? Show me what happens when tests arrive on your door step. Then team goes to school unannounced and has set of particular **questions that are used there. How did you train staff? Etc.

If anything is found during the assessment visit, the school is given the opportunity to report that to the Office of Management Assistance. One way or another it gets reported.

- 2. The State Department of Education must provide training of test administration, especially test security.**

TRAINING OF TEST ADMINISTRATION

In collaboration with the Office of Assessment and Accountability (OAA), the Division of Assessment Implementation (DAI), and the Office of Exceptional Children, training is provided annually to all DACs and directors of special education in every school district. Everyone is expected to attend. Anyone involved with testing is responsible for appropriate implementation of the state testing program. Those who receive the training are responsible for training anyone in their district who will have anything to do with testing in the district.

In addition certified staff of the schools have a professional code of ethics that they adhere to. Any violations are reported to the professional standards board.

- 3. The State Department of Education must have a policy on repercussions for educators who breach test security.**

BREACHES

Allegations may come from any source. There rarely have been direct violations found during the random monitoring visits to schools and districts during the testing window. However, discussions have been held at the school/district about how to improve the appropriate testing practices.

In the administration code the process for allegations and investigation of allegations is laid out. A formal investigation is pursued. All results go to the board of review made up of department staff who are not employed in the OAA. After reviewing all the details, the board of review makes recommendations to the commission. These recommendations may include: no action, score changes, action taken on the individual involved, recommend that it be turned over to the professional standards board, recommend a more thorough investigation or even a full scale audit.

(If there is an allegation, the first question asked will be “did you attend the above training.”)

For perspective, in April, 2000 a total of 463,360 students were tested. There were 184 allegations or investigations, of which 62 were substantiated incidents. The 1,081 students with changed scores resulting from those investigations constitute 0.2% of those tested. Many of these grade changes represent errors in judgment, not intentional cheating.

If any student's score is changed, the score is only changed for accountability purposes at the school level. The score is not changed at the student level.

The responsible party is the test administrator.

For the Administration Code

http://www.kde.state.ky.us/oa/Implement/Adm_Code/Adm_Code_default.asp

Attachments

COMMONWEALTH ACCOUNTABILITY TEST SYSTEM

Kentucky Core Content Test-Spring 2001
Appropriate Assessment Practices Certification Form

NONDISCLOSURE AGREEMENT

References

COMMONWEALTH ACCOUNTABILITY TEST SYSTEM

Kentucky Core Content Test-Spring 2001
Appropriate Assessment Practices Certification Form

Appropriate Assessment Practices

The Kentucky General Assembly through KRS 158.6453 requires an innovative student assessment program designed to measure student progress toward achievement of the goals specified in KRS 158.6451. The purpose of this document is to describe the practices that are considered appropriate in preparing students for the assessments, in administering them, and in assuring proper security of the assessments. In determining appropriate practices, the following standards have been used: ¹

- Professional Ethics No test preparation practice should violate the ethical standards of the education profession in 704 KAR 20:680.

- Educational Defensibility No test preparation practice shall increase students' test scores on the state-required tests without simultaneously increasing students' real ability to apply the content to real life or simulated real-life situations.
- Student Ownership All assessment work shall be done entirely by student. No one shall coach, edit, or point out errors in student work on the open-response or multiple-choice portions of the test.

The first standard simply refers to evaluating the appropriateness of a practice to determine whether it meets standards of professional ethics. Therefore, such activities as duplicating secure testing materials, which is specifically prohibited, would be considered inappropriate by applying this standard. Only those items designated by the department or its contractors as "released" may be reproduced for any reason without the specific written permission from the Kentucky Department of Education.

The second and third standards encourage any activities that increase student mastery of the content being tested, but would classify as unacceptable any practice that creates a deceptive picture of student achievement.

The Kentucky Core Content Test is a standardized assessment consisting of open-response and multiple-choice items and on-demand writing prompts. This type of testing is generally familiar to most teachers, and the practices that are appropriate are well known. It is appropriate for teachers to know the concepts measured by the state-required tests and to teach those concepts. Teachers may use test items from previous years released by the Kentucky Department of Education to prepare their students for the testing. On the other hand, it is unethical to know specific test questions before the administration of the test and, more importantly, to use the knowledge about the content of any specific test questions in preparing any instructional materials or delivering any lessons. Administrators and teachers should use all due diligence to ensure the security of the test before, during, and after administration. All copies of materials shall be returned to the testing contractor immediately following administration or properly secured through procedures that may be issued by the department. No assessment materials other than those officially designated as released may be duplicated. All directions in the *Instruction Manual for District Assessment Coordinators and Building Assessment Coordinators* and the *Administration Manual for Test Administrators and Proctors* that accompany the test materials shall be followed. For additional information concerning appropriate assessment practices, please refer to the *Administration Code for Kentucky's Educational Assessment Program*.

All Spring 2001 Appropriate Assessment Practices Certification Form signature block is provided on the back of this page. **All** school personnel involved in the assessment must sign this form. Additionally, an administrator/proctor guide for the state-required assessments is provided for use by test administrators.

¹Popham James, Appropriateness of Teachers' Test-Preparation Practices; *Education Measurement: Issues and Practices*, Winter, 1991

Administrator/Proctor Guide for the Kentucky Core Content Test

Administrators/proctors for the Kentucky Core Content Test shall maintain an atmosphere conducive to the successful completion of the assessment. No information about the content of answers may be provided to the students. These recommendations are made in the spirit of fairness such that no assistance is rendered to some students that is not offered to all students.

Administrators/Proctors may:

- Explain to students how to fill out test booklets according to the directions
- Provide extra writing instruments, sharpeners, highlighters, blank writing paper, blank graph paper
- Circulate among students during testing
- Prevent observers from distracting students
- Verify beforehand that the testing environment will not prohibit students from completing their tasks (e.g., adequate ventilation, no broken furniture, adequate lighting)
- Take all measures allowed by the school to prevent interruptions (e.g., messengers, intercom announcements)
- Cite and attempt to correct printing errors if they are discovered (e.g., miss-numbered pages, blank pages, smeared copy)
Notify your District Assessment Coordinator if you believe it is a universal or widespread error.
- Correct the spelling of misspelled words in the test booklet if necessary. The only spelling errors worth correcting are those that would cause a change in the response
- Accommodate for physical comfort as necessary (e.g., broken chairs, drafts, lighting)
- Encourage students to work constructively to complete the assessments; however, proctors shall not provide any content information for the assessment

Administrators/Proctors may not:

- Read the questions to students (except as specified in the manual)
- Provide references other than specifically indicated in the instructions
- Provide tools except as specifically allowed in the instructions (i.e., calculators shall be provided for mathematics and science and may be used in other content areas; dictionaries and thesauri may be provided for writing prompt **ONLY**)
- Answer questions related to the response (no hints, restatements, interpretations, rephrasing for clarification)
- Threaten students or lie about consequences of testing performance
- Provide preprinted acronym sheets or paper containing a system for organizing answers (e.g., column method, ROOTs)

Administrators/Proctors should not:

- Hover over individual students for extended periods of time
- Allow distractions
- Engage in activities preventing their full attention to the students

All school personnel involved with the assessment program must sign the following certification form. The form should be kept on file in the local district.

COMMONWEALTH ACCOUNTABILITY TESTING SYSTEM

KENTUCKY CORE CONTENT TEST-SPRING 2001

APPROPRIATE ASSESSMENT PRACTICES CERTIFICATION FORM

I have received and read the *Administration Code for Kentucky's Educational Assessment Program* and the *Appropriate Assessment Practices Certification Form* provided by the Kentucky Department of Education.

Name of District

Name of School

Staff Member's Signature

Date

**COMMONWEALTH ACCOUNTABILITY TESTING SYSTEM
(CATS)**

NONDISCLOSURE AGREEMENT

The Commonwealth Accountability Testing System (CATS) student assessment is a project of the Kentucky Department of Education. The design of the program requires that the test questions remain secure. To protect the security of the tests, only authorized persons are permitted to view the test questions. With the exception of items released by the Department with official CATS reports, all CATS test questions, draft or final, and all supporting assessment materials or notes are to be regarded as secure documents. Thus, they may not be reproduced, discussed, or in any way released or distributed to unauthorized personnel. Any reports or other communications based on the authorized use of secure CATS materials must be reviewed by the Office of Assessment and

Accountability or its designated agent prior to distribution or dissemination by the user. Permission of authorized use of secure CATS materials may be granted by the office, provided the user agrees to abide by the terms stated herein.

The undersigned is an employee, contractor, consultant, advisory committee member for the COMMONWEALTH ACCOUNTYABILITY TESTING SYSTEM (CATS), or person otherwise authorized to view secure CATS materials and hereby agrees to be bound to the terms of this agreement restricting the disclosure of said materials.

Name (printed)

Signature

Date

References

Administration Code

http://www.kde.state.ky.us/oaa/implement/Adm_Code/Adm_Code_default.asp

Kentucky Department of Education. (2000). *District Assessment Coordinator Implementation Guide for the Commonwealth Accountability Testing System*. Frankfort, KY. Author.

Kentucky Department of Education. (2000). *Administration Manual for Test Administrators and Proctors*. Frankfort, KY. Author.

Discussant's Comments

Joe O'Reilly

Mesa Public Schools

This session has lots of good news. First, there isn't a lot of cheating going on out there –

- ❖ Greg cites a teacher survey that found most teachers rated cheating as rare or never.
- ❖ Mary found that test directors also said it was rare. And two-thirds of her respondents didn't have a good cheating anecdote to share.
- ❖ Linda found 62 substantiated instances of cheating out of 463,360 students tested. While every instance of cheating surely was not reported, this is evidence of the low rates of cheating.
- ❖ Michael also believes the instances of cheating are much lower than the media reports.

But I think people believe there is a lot of cheating out there because they have heard about some cheating and one story goes a long way. In fact, there is a whole subsection of urban legends on cheating. Here is just one --

Figure 1 Cheating Urban Legend

Legend: A student stops by the office of one of his instructors to ask a question and finds that the professor has stepped out for a moment, leaving an unguarded stack of the next day's final examinations on his desk. The student quickly steals one of the exams and disappears. Issuing the exam, however, the professor counts them and notices that one is missing. He cuts a half-inch off the bottom of every exam prior to distributing them to the class, then fails the student who turns in a test paper longer than the rest.

Origins: Many collegiate legends deal with students who devise clever ways of cheating by fooling teachers with switched examination papers and booklets (such as *Book of Daze*, *Paginal Exam*, and *Mother Knows Test*). This example presents a reverse twist on those legends: the cheater is outsmarted, caught, and punished by a quick-thinking instructor.

Source: <http://www.snopes2.com/>

These urban legends also show that cheating has been around for a long time. Why then is it such a hot topic to local test directors now? It's because of the increased focus on testing. I have to tell you, when I watched the presidential debate and the candidates, who usually don't even mention education, started arguing over who was going to require the most testing I turned to my wife and said, "Oh no (or a word something like that) – this is not a good thing! I think there is an old saying that goes something like 'when the gods fight, the people suffer' and we are the people.

It is also a hot topic because, as Michael so aptly put it, 'it's the stakes!' In fact, with an increase in high standards graduation tests, tests determining continued supplemental school funding, proposed school takeovers based on low scores, and other high stakes outcomes we are seeing the culmination of pressures not seen before. To take a page from our meteorologist friends, the conditions are making it probable that we are approaching a 'Perfect Storm' of cheating.

As our presenters have pointed out, adults tend to cheat when they see tests as being high stakes for them or if they see it as inappropriate or irrelevant to them and their students and the results are not useful. Students tend to cheat when the stakes are high and when the testing situation is unfair.

When the stakes are high for just the student, the teacher tends to catch, or at least try to catch, the students who cheat. If the stakes are high just for the teacher, the student or parent or even other teachers often turns in the offending educator.

But when it becomes a 'win-win' situation for both parties, as is happening with some of these high stakes graduation tests, we have a convergence that looks very troubling to me.

So how well prepared are we for cheating? Mary's study gives us more good news ---

- ❖ Standards are generally prevalent in school districts.
- ❖ Actions consistent with the Standards were rated as very important or extremely important by test directors in 22 of 24 instances. And, although our self-reported actions don't always match our desires, in practical terms they were fairly close.

I should note that I agree with the other test directors that one of the 24 actions is not important – the superintendent publically reporting breaches of test security. In the media world, where 'if it bleeds it leads,' this could have its own negative outcomes – painting the other 95-99 percent who don't cheat with a black brush.

I can't think of any good reason for a superintendent to engage in self-flagellation in full sight of the general public. However, I do think that most test directors would agree that it is very important that violations of test security be very publically addressed within the school system. We make a concentrated effort to let teachers and

principals know that we have caught people and they have suffered consequences. That warns the tempted and lets the honest folks know the system is fair. But we make an equally concentrated effort to avoid public shaming of individuals and giving the public the impression that cheating is widespread.

So, while the test directors think that the Standards are important, and generally act in accordance with them, is that enough? Probably not, because the test director is the only person bound by the Standards. Systematic rules for testing must be in place. Greg's paper offers a list of steps that test directors should consider. For example, we produce a "Directives for Testing" that teachers must sign that they have read and we have test monitors randomly visit classrooms during testing, two actions included in the list of suggestions. Another suggestion, instituting an honor code, is something I am now looking in to.

I should note that some of suggestions, such as changing test formats from multiple-choice to essays, adding bar coded booklets and delaying delivery of tests may not be administratively practical when you have to process literally hundreds of thousands of tests in a six week period and a teacher/test coordinator has to fit in counting out tests while still keeping her 'day job.' And I am afraid that the suggestion of very strong, 'third rail' type penalties such as a life time bar from teaching may go too far and actually reduce the chances that a colleague or a supervisor would turn a teacher in for cheating.

Greg's paper also discusses the ways we catch cheaters – unusually high levels of performance, statistical analyses, etc. You didn't really dwell on my top two resources for catching cheaters – colleagues and parents who are teachers in another school. If something odd is going on, these are the people most likely to tell me. My next biggest source is eyeballing the bubble sheets. If a teachers' class has a lot of erasures we check the percent going from wrong to right and vice versa.

What was notable to me was the lack of good available tools for catching cheating like this. We hand do our ‘erasure analysis,’ but publish it in a form that looks like a computer printout so it would be more believable. [I haven’t had anyone challenge ‘the computer’ yet.] What I would really like to see is a program that actually does an erasure analysis.

Based on my reading of these papers, and the fact that this is the NATD symposium, I would like to suggest a few steps that NATD could take to help test directors prepare for the possibility of a ‘perfect storm’ on the horizon.

- ❖ Develop a model code for teachers and administrators on how to prepare for, administer, and score tests. This shouldn’t be a collection of samples from districts, but an actual model code put together by an ad hoc task force that can be taken by a district, tweaked and put in place.
- ❖ Develop a similar model code for students.
- ❖ Provide sample forms that can be used to have teachers and students acknowledge these codes.
- ❖ Provide sources for catching cheaters – as many resources as possible from a review of Scrutiny and other products to a simple table listing the probabilities of going from wrong to right and right to wrong on a multiple-choice test.
- ❖ Develop a ‘best practices’ standard for what is the appropriate assessment training for principals and teachers called for by Michael.

There is something else I would like to see, but I am not sure if NATD is the appropriate group to ask -- the FBI or National Institutes for Justice may be more appropriate. I would like to see some advice on how to handle the nuts and bolts of a cheating investigation – how to interrogate the accused, what evidence should be collected and how much is enough, etc. My personal pet peeve is the teacher who says, “but the rules didn’t say I could . . . “[insert a ridiculous answer here, such as “I

didn't know we couldn't tell them they had wrong answers and to try again, I was just trying to help them do their best!], but that is to be expected in a country where the coffee cups come with warning labels that the coffee may be hot.

Finally, I would like to raise two questions. First, I would like to pose a question to Michael that he can answer in the Q & A section – please describe the parent/educator/community committees that would supervise assessment efforts – how would they work and what major benefits do you think they would provide? Second, I would like to ask the panel how they think technology (sharing information on web sites, wireless communication devices, etc.) will impact us and how we can/will respond.

So to sum up, this is not a good news topic, but this is a good news session. There is some cheating going on, but it is not epidemic, at least not now. And there are a number of steps we can take --making the results more useful to better proctoring that can help keep this problem to a minimum. I think this has been a very valuable session from the practical point of view of a test director.