Handbook of Test Security

James A. Wollack, John J. Fremer

Security Issues in Technology-Based Testing

Publication details

David Foster
Published online on: 26 Mar 2013

How to cite :- David Foster. 26 Mar 2013, Security Issues in Technology-Based Testing from: Handbook of Test Security Routledge
Accessed on: 21 Dec 2018

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Introduction

Tests have been administered on computers in a large-scale way since the late 1980s. One of the first high-stakes programs to do so was the licensure test for the National Association of Securities Dealers, using the PLATO system (Wikipedia, 2009). Since then, many, if not most, large-scale testing programs outside of education have converted paper-and-pencil programs to ones using computers and other technology. Testing programs starting from scratch at this time generally begin with the intention to administer tests via computers. There is no standard or obligation to start a high-stakes testing program first with a paper-and-pencil test. Despite the advances and popularity of technology-based testing there remains a large segment of testing programs, mainly in the area of education, that continue to use paper-and-pencil test administration. This chapter will focus on technology-based tests and their security issues. Other chapters in this Handbook (Chapters 2 and 4) provide similar information for paper-and-pencil tests.

Terminology

Before proceeding, it is important to clarify terminology and definitions. These terms follow no hard-and-fast rules and generally refer to the way tests are administered, rather than to the use of technology throughout the various steps of the process, including development of items and tests, collection of results, conducting statistical analysis, and reporting. Paper materials are often also used in tests that mainly use technology. It is a rare and disciplined technology-based program that never prints out a paper version of items to review, or a report. Similarly, paper-and-pencil tests will use computers and the Internet often as needed. Clear terms, however, are important for categorizing distinct categories of technology use.
• **Paper-and-Pencil Testing.** Paper-and-pencil testing is a term referring to the traditional method of giving a test that is made of paper. It usually includes the use of machine-scorable answer sheets. The tests themselves may have been created and reviewed on paper, although it is fairly common today to create paper-and-pencil tests with the assistance of the computer or by using the Internet, and then print them out for administration.

• **Computerized or Computer-Based Testing (CBT).** In this chapter, computerized or computer-based testing will be used interchangeably and refer to the use of computers (including desktops, laptops, and even mobile computing devices such as tablets and smartphones) predominantly to administer tests, although it will include all phases of test production, administration, results management, analysis, and reporting. The terms are restricted in their use to mainframe computers, networked microcomputers, or stand-alone computers. In this chapter, the terms will not refer to the use of technology for test administration that is completely or mostly Internet-based.

• **Computerized Adaptive Testing (CAT).** A term often used by testing professionals when referring to computerized testing or CBT, as if the two terms were synonymous, probably because many of the original computerized tests were CATs. A CAT is a particular test design that selects questions during an examinee’s test depending on the previous responses of the examinee.

• **Internet Testing or Internet-Based Testing.** These terms refer to testing where the Internet is the dominant technology for test administration. The Internet, as a new technology with many of its own security issues, brings a new approach to testing justifying a separate category from computerized testing or CBT.

• **Technology-Based Testing (TBT).** Often in this chapter, to accommodate the convenience of being more general, it will be necessary to refer to all of the testing models, with the exception of paper-and-pencil testing, as technology-based testing.

**Purposes of this Chapter**

This chapter will describe the security issues that arise when using technology to create or deliver high-stakes tests. A good definition of high-stakes testing is provided by the Association of Test Publishers (2002, 3):

> High-stakes tests are situations where decisions and interpretations from test scores have important, critical, and direct consequences for the test-taker, the program, and the institutions involved. Examples of high-stakes testing environments include final examinations for a college course, college and graduate school admissions examinations, professional certification and licensure tests, and job selection tests.

The chapter covers what should be done to prevent and deter security problems, or, failing those, to detect and mitigate them.

The chapter will not cover security issues that are solely the realm of paper-and-pencil testing. For example, teachers may help administer exams for state assessment to their students. When the test is completed, the teacher may change the answer sheets by erasing incorrect answers by his or her students and replacing them with...
correct answers. This security problem, and similar problems that are unique to paper-and-pencil, is well-documented and is covered elsewhere in this volume.

One additional prefatory comment needs to be made about the prevalence of security problems in all areas where test results have important consequences, namely high-stakes testing. Security problems have always existed and always will. This is true for all types of testing, whether paper-based or technology-based. And cheating occurs in all fields, including education, certification, licensure, clinical psychology, and industrial/organizational psychology, and also extends beyond testing (e.g., sports, investments, taxes, gambling, etc.). It is equally true that, as individuals, testing programs, and society as a whole, we have not tolerated test fraud of any type, and have always responded with solutions. The use of new technology to administer exams provides new challenges which take time to discover and solve. It is important to know that tried-and-true security methods for paper-and-pencil testing often do not work as well for technology-based tests. Nevertheless, we remain convinced of the danger test fraud poses to the credibility of the testing process and to the measurement industry as a whole. Likewise, if unrecognized or ignored, which has happened in some circumstances, the damage continues to occur.

While it is true that we will never be able to stop all test fraud, it is equally true that we can always minimize its damaging effects and reduce how often it happens. We can do this by the motivated and intelligent application of security efforts. Much can be prevented or deterred; all types of security threats can be detected and dealt with. It is the responsibility and should be the goal of all well-managed TBT programs to reduce the influence of cheating and other security problems on test scores, and on the important decisions those scores support. It is the right of every test taker to obtain a score that is fairly achieved in an environment of secure testing.

In addition to describing the threats produced by the adoption of technology in testing, this chapter provides a framework for risk management by categorizing those threats and describing solutions for preventing them and reducing their effects.

**Reasons for Security Problems Specific to Technology-Based Testing**

With the use of new technologies to assist the testing effort, new security threats have surfaced. Even using a modem and phone lines, instead of a postal service, to transmit test files creates the chance that a “hacker” will intercept the file being transmitted. Obviously that would not be possible if an electronic transfer method were not used. Of course, the threat of theft should not stop our use of such technology, but instead motivate us to anticipate the security risks and either prevent them or detect them when they happen and deal with them.

There are several general reasons why technology use in testing has produced particular new threats with which we, as an industry, have been historically unfamiliar.

**On-Demand Testing**

One of the major advantages attributed to TBT is the convenience of being able to take a test “on-demand.” While most technology-based exams are not available for strictly immediate testing, “on-demand” generally means that you can schedule a test at a more convenient time for you than would be available by the more typical scheduling
used for the paper-and-pencil testing model. There is no need to choose one of the few required testing sessions provided; instead, you can choose any available time slot in a testing window that may be a few days in length to a few weeks or months. Some tests have no long-term limit to that window.

Why does on-demand testing increase the risk of security problems? Testing windows leave test content vulnerable to theft and encourage cheating. There are test takers who take the exams earlier in the window, and there are those that take it later. In the simplest of examples, an earlier test taker sees the content of the exam while taking it. Then afterward he or she informs later test takers of that content. While the first person may not be cheating for personal gain, he or she is stealing the content and colluding with others. The later test takers, armed with pre-knowledge about the test, cheat by using that information to get higher scores on the test. The use of on-demand testing also attracts more organized efforts. In a well-publicized incident, in 1994 Kaplan Test Prep sent test takers to take the new computerized GRE exam published by ETS, asking them to memorize sections of the exam questions. After completing the tests the individuals were debriefed, allowing Kaplan to capture as much of the new test as possible. As another and even more devastating example, in the information technology certification industry, test questions are stolen and sold almost simultaneously with the test’s release (Foster & Zervos, 2006). The thieves do this by colluding with a testing center, capturing test files that have been previously downloaded – for a scheduled exam event – to the center’s servers. Foster & Zervos purchased hundreds of tests from braindump sites and compared stolen items with the original items. The result for an actual exam is shown in Figure 3.1. A score of 1.0 indicates that 100 percent of the “words” are found in both texts. Here is Maynes’s description of the match analysis algorithm:

A word is a string of alpha-numeric characters. Punctuation is discarded. Case is discarded. Order is discarded. Answer-option labels are discarded. Hyphenated words are turned into two words. Percent symbols in the ends of words are converted into the word “percent.” Punctuation characters are: single quote, double quote, period, comma, colon, semi-colon, question mark, exclamation point, percent sign, left parenthesis and right parenthesis.

(Maynes, personal communication, 2012)

In Figure 3.1, for almost 120 items, each pair of items shared between 95 and 100 percent of the words. Only two items matched at levels less than .95.

Stolen questions are provided on so-called braindump sites as a “test preparation” product. While these sites are often called braindump sites, referring to the act of test takers “dumping” the test content in their brains to these sites, it is true that test file theft, and not memorization, is the main method of capturing this test content.

The longer the testing window, the more opportunities there are to capture and share test content. When the window is very long, even indefinite, the availability of the exact questions remains useful to a large number of later test takers, and continuously reduces the value of the exam. Recent research (Maynes, 2009) has indicated that, for several large IT certification programs and particularly high-volume exams, a majority of examinees may be using questions/answers purchased from braindump sites. I will return to this later in the chapter.
Mostly due to this particular security problem, the information technology (IT) certification industry suffers from a serious problem of distrust and unfairness. Honest certification candidates do not trust programs that allow such theft to happen and do not like the fact that users of braindump sites are allowed to prosper. The programs themselves cannot be sure which candidates used the braindump sites and which did not. All scores are suspect as are the certification decisions based on them. And certainly organizations that depend on certified individuals for critical positions within the organization are unable to fully trust those that come with certification in hand.

Content that is stolen is shared in many ways. As described above, a common and effective method is simply to post the entire set of questions for sale on a braindump site. Such sites can easily be found by doing a web search using the terms “IT certification” and “actual test questions.” However, more casually, test takers can simply recall the questions seen and send the content of the questions through email or text messages, or discuss them openly on chat rooms or forums. They can also be compiled into test preparation resource books and distributed with reference textbooks. The longer the testing window, the more useful are each of these methods to the cheater.

Reducing the size of the testing window is a partial solution, but mostly ineffective. The ability to steal items and communicate them to others usually occurs in electronic ways and is very efficient with today’s social networks. Even in a supposed “simultaneous” test event that is limited to the mainland of the US, items can be stolen on the east coast of the US and provided to test takers taking the same exam at the same hour of the day on the west coast. Removing or reducing the testing window also defeats the original advantage of convenience to the test taker.

**Electronic Distribution of Test Files**

Electronic distribution of test files leads to quick and easy technological ways of capturing the content. Since the late 1980s and continuing to the present time, computerized tests in the form of test files are typically sent from a server residing at the offices of the testing program (or a test administration contractor of the program)
to a server at a testing center. All the files, including the test questions, scoring rules, answer keys, relevant graphical, audio and video files, and other components of the test, are usually encrypted and included. The files may be sent a few hours to a few days before the particular test is scheduled to be taken. It is at this step during transit and when the files reside on a testing center’s server that they are vulnerable to capture (see Foster & Zervos, 2006). Pirates often collude with a testing center to access the testing center server, capture the file, decrypt it if necessary, and put it into a format that can be easily sold on the Internet. This type of theft is particularly devastating because the complete pool of items supporting the forms for an operational exam is stolen.

**Storage of Item Banks, Tests, and Other Information**

Development and storage of item banks, test files, test taker information, and test results on a program’s servers provide new opportunities for technically skilled individuals to obtain that information. As tests are created today, most of the steps the items and tests proceed through are in electronic or digital format. Data are stored on a program’s servers whether controlled by the program or hosted by a service. Managing access to those servers is the responsibility of the testing program. However many testing programs do not have the experience or knowledge to manage user access effectively. Employees and contractors, including administrators, item writers, reviewers, psychometricians, and many others, have access to most if not all of the items, tests, or other important information.

With the growing sophistication of database systems and capable hackers who have the goal of penetrating such systems, most testing programs find it difficult to keep up. They often do not have up-to-date information security systems or sufficient control over access to the testing data to prevent attacks. Even employees or contractors who have left the employ of the organizations often retain access rights to the testing databases long after they have gone. Protecting testing data certainly has to start at the source. Every TBT program has to give high priority to protecting the data that exists under its direct control.

**Test Question Overexposure**

New technology-enabled test designs such as CAT increase exposure rates inordinately for some questions (Folk & Smith, 2002). Computerized adaptive tests select questions from the same pool of items when each new examinee takes the test. This results in both security advantages and disadvantages. The disadvantage is that adaptive tests use algorithms that select the “best” questions to present to a test taker at any particular ability level. These questions are considered best by the program because they provide the most information about the test taker’s ability when answered. If a question is psychometrically better than another for a set of students, it may be selected more often than other questions and therefore becomes overexposed compared to other questions. The increased frequency becomes apparent and such questions are then stolen and shared more often in casual or Internet discussions as a result. Such overexposure may lead to disclosure and to unfair advantage for those test takers who have access to information about the question. Any overexposure of
these items is particularly destructive because the items affected, by definition, are the best, most useful items.

Test designs that use a cutscore to determine a pass/fail decision may also overexpose questions unnecessarily. With technology-based tests it is possible to determine that the test taker has answered enough questions to pass or fail the test. At the point where the pass/fail decision is certain, there is simply no reason to continue to present questions to the test taker. Every instance of item presentation is an instance of overexposure.

Most TBT engines have incorporated a feature of marking an item for later review. This is similar to what a test taker is able to do if given a paper-and-pencil test, which is to move from page to page at will and review any question as often as desired. The security risk lies in the fact that the test taker is able to track and control a particular subset of questions which he or she can memorize in a more organized and efficient way, say, at the end of an exam. If these questions are reviewed and memorized at the end of an exam it is much easier to leave the testing center and recall the items.

**Testing Centers**

The testing center network model provides new challenges for security, particularly those centers that are “franchised” or “authorized,” that are owned and operated by an organization other than the testing program or its test administration contractor. With the advent of computerized testing in the late 1980s and early 1990s came the testing center model for technology-based tests. The first large-scale world-wide introduction began in 1990. Novell, Inc. had a growing international certification program in the area of information technology, but it did not have locations where tests could be taken in a standardized way and with security measures in place. It contracted with Drake Training and Technologies, the original Prometric organization, to deliver the exams and it encouraged the contractor to use Novell’s more than 2,000 IT technology training center locations. That decision placed those testing centers under the direction of its training partners. In a fairly short time period each training center had a testing center as well. This particular franchised model still exists today and serves as a way for most IT certification, and many non-IT, testing programs to distribute tests internationally. Today, there are several similar networks in place with a total of many thousands of locations. Combined they covered the United States very well, and other countries more sparsely.

While it was necessary at the time to use training centers as testing centers, the arrangement unfortunately contained an inherent conflict of interest: training companies had a business interest in how well their students or others performed on the tests. The situation caused some testing centers to use instructors as proctors, helping out during the test when needed. At the most extreme level, the worst testing centers actually colluded in many aspects of test theft, including allowing cheating, proxy test taking, and stealing the test content (Maynes, 2009).

Another security problem with the testing center model, particularly the franchised model, lies with the proctors. In general these individuals are lower-paid, are part-time, and have received minimal or even no training. These proctors are generally unmotivated to closely monitor test takers and detect cheating or theft of questions. Many recent public breaches have been because proctors have
assisted in the dishonest behavior. Even if the proctor is honest, lack of training and motivation make it difficult to confront a person who is suspected of cheating and deal effectively with the problem. This problem with proctoring is not specific to testing center channels where technology-based tests are administered, but has been a long-standing issue as well in large-scale educational testing with paper-and-pencil tests (Cizek, 1999).

Finally, authentication of the test taker’s identity at testing centers is usually accomplished by showing a government-issued ID. With new technology aiding in the creating of very good fake IDs, it is very difficult for center administrators or proctors to recognize the difference. This situation would allow individuals to take tests for others even when the proctor follows proper procedures.

**Internet Test Administration**

The benefits of the Internet have encouraged testing programs to administer high-stakes tests widely across the Internet without proper security in place. This is usually done to extend the reach of the exam, providing the ability to easily attract and impact a very large number of individuals. For example, some organizations specializing in pre-employment screening tests have provided these exams almost indiscriminately over the web. The ease of distribution has encouraged organizations to administer tests over the Internet with no supervision or ability to authenticate the test taker (Tippens et al., 2006; Foster, 2009). These tests have important employment consequences for the participant because they are used to screen applicants for a small number of jobs. Those applicants with higher test scores move to the next level in the hiring process. There is evidence that these tests are being stolen, cheating is happening, and the lack of security is being exploited.

**Specific Security Risks with Technology-Based Tests**

It is clear that the introduction of technology for test administration has also changed the nature of the security risks. It is important that the reader understand the risks associated with technology in order to choose between and implement the solutions that are available. These are presented in a later section. No field of high-stakes testing is more immune than another when considering these risks. Areas of testing where examples are provided in this chapter are: certification; licensure; industrial/organization psychology; K-12 and higher education; distance education; and college admissions.

Using technology has not introduced any new overall category of security problems. **Piracy** and **cheating** remain the two main categories of security problems whether the tests are paper-based or technology-based. Cheating is defined for this chapter as any effort that produces a score higher than what is deserved either for oneself or others. Test theft is defined as any behavior that inappropriately or illegally captures test questions and/or answers. Test theft occurs for the purpose of sharing with or selling to others who will then use the information to cheat.

It is important to note that cheating is not necessarily done by the test taker, or with his or her awareness. Many cheating incidents occur and are managed by others, particularly in education. These people may be teachers, principals, or other officials.
In addition, many test takers may be unaware that they are using pre-knowledge of actual test questions or that using such pre-knowledge is actually cheating.

It is useful to subdivide the two categories of cheating and theft into subcategories of general methods of cheating and test theft. I have done so in Tables 3.1 and 3.2.

**Test Theft**

Table 3.1 ranks seven different types of test theft. In each case, the intent of the theft is to capture the intellectual property, or test questions, of a testing program. The categories are ranked from most serious (1) to least serious (7) loosely based on ease of theft, completeness, opportunity, potential damage to a program, and frequency.

**Recent Examples of Test Theft**

Stealing Complete Test Files: IT Certification

Some systems can be easily hacked in order to access item pools. With ever-advancing technology and the advent of the Internet, it has become increasingly difficult for testing programs to keep up with the changes in IT security – or more appropriately, with the ever increasing security risks associated with this new technology. New operating systems seem to open more security holes than they fill. Browsers are becoming more friendly and capable, allowing access to web applications and their databases directly if special care is not taken to manage them.

As an example of one of many IT security risks, many applications have authentication and session management systems that are either mismanaged or entirely too simple. Such systems can be undermined by flawed handling of password changes, “forgot my password” and “remember my password” functions, account updates, and other related functions. Many systems do not even require reauthentication after a period of time has passed with no activity. These flaws, and others, can result in the compromise of user or system administration accounts, allowing an attacker to hijack an active session and assume the identity of the user. If that user were a testing program manager or a system administrator, the attacker would have easy access to all of the item banks on the servers.

Maynes (2009) and Foster & Zervos (2006) describe the theft of test questions from files that have been downloaded to testing centers. Foster & Zervos purchased hundreds of test files in Word documents from an IT certification “test preparation” website – www.certexperts.com – which was going out of business. The entire package cost a mere $200. After examining the files, it was discovered that they contained more than 60,000 questions for certification tests for nearly 60 different IT certification programs. Furthermore, the files contained the exact text of the questions (see Figure 3.1 for an example from one program), even including minor punctuation nuances, and the order of the questions in the files was the exact order of the files in the original published files. The significance of the latter fact is simply that it led the researchers to discover the mechanism of the theft. Since test questions are routinely administered randomly, the order on the Word files would have been more or less jumbled. Because the items were in the same order as the source files, it is clear that the theft did not occur during the exam itself, but before or after the exam while the original pool of
<table>
<thead>
<tr>
<th>Brief description</th>
<th>Reason for ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Stealing actual test files</td>
<td>This method is the most dangerous type of theft because it provides the exact content of the questions and the answers. Once the method of theft is established, stealing any test file is routine, with virtually no risk of detection.</td>
</tr>
<tr>
<td>Test takers memorize questions to be recalled later</td>
<td>The method needs no technology, but is difficult to do for tests with a large number of items or where items are pulled from a pool. It also results in inaccuracies in the item content. No answer key is captured. Probability of detection is low.</td>
</tr>
<tr>
<td>5. Violating retake rules to gain additional access to questions</td>
<td>This generally costs the thieves additional testing fees. Because the rules allow for legitimate retests, the chance of catching someone who exploits this for purposes of item harvesting is very low.</td>
</tr>
<tr>
<td>7. Obtain test material from program insider</td>
<td>The content will be accurate and damaging to a program. This type of fraud is rare and the risk of detection and criminal punishment is high.</td>
</tr>
</tbody>
</table>
Security Issues in Technology-Based Testing

test questions resided on the testing center’s server. Although this type of theft could also implicate an insider at the testing program or the program’s test administration vendor, neither was possible in this case because insiders did not have access to the complete set of items for all the programs included in the purchased package.

The Word files had been created from one or more files containing the test questions, answers, and supporting resources (e.g., graphics). Further review of the files showed all graphics were embedded with the corresponding questions, including those graphics used for more interactive questions, such as drag-and-drop and point-and-click. The areas on those graphics where a test taker would need to drag and drop an object, or simply point and click, had been clearly marked. While the Word files could not “show” how these answers were to be produced, it did show the correct results. The files also contained the answer to each question, following the question. It was obvious that, after the original files were stolen, some minimal manipulation of the files occurred to put them in a document for easy use by purchasers.

So, how might the thieves get access to such test files? First, the files for any particular test are downloaded to a testing center anywhere in the world when a test is scheduled. The files are downloaded some hours or even days before the test is to begin. Proper security at testing centers would have these files protected in two main ways: system access rights and file encryption. System access rights would only allow trusted testing center personnel access to areas on the server where the downloaded test files are stored, and file encryption would help make sure that, even if the files were captured, they could not be decrypted and used. Still, one could gain access to this information if the system had only minimal system access controls. That is, either there is no system access control, or the control system is so simple that finding the usernames and passwords is easy to do. Another way one might gain access is through the test administrators at the testing center. They may be incompetent, setting up usernames and passwords that are too simple, or untrustworthy, simply being too free with the access information to other employees of the testing center. Finally, it is possible that the testing center administrator is actively assisting with the theft. He or she may be part of the scheme to quickly capture test files, decrypt them, and put them in a format for sale on the web. Most testing centers throughout the world are labeled as “authorized.” They generally are part of another business and are franchised by the test administration vendor. Quite often these centers are part of a training operation, usually related to the certification programs for which the tests are administered. This inherent conflict of interest may encourage a small percentage of the many thousands of testing centers to steal and sell test questions, both for the business value and as a “service” to their students who are training on the subject matter. Because these are IT training organizations, it is not unreasonable to assume that a few of the many thousands have the technical ability, opportunity, and motivation to steal and sell the questions. Within those centers, it often only takes one individual who is willing to compromise his or her ethical standards for there to be a security breach.

Visiting any braindump site, one finds the blatant statements that they have the actual test questions and that purchasing their test preparation files will result in a guaranteed passing score. Both of these claims are frequently true, unfortunately, and the effect is devastating to IT certification programs. By association, any area of certification suffers from this cancer, even if a program’s questions have not been stolen and are
not offered for sale on braindump sites. And it is not only the certification program that suffers. Candidates for certification do not obtain a score that can be trusted for making certification decisions. The certification, which should have significant value in our society at this time, is tainted and viewed with suspicion. The biggest losers, however, may be the organizations that hire these certified individuals, expecting competence. What they find is that they have allowed incompetent individuals access to very important mission-critical systems and applications. These individuals, who must then learn on the job, may do much more harm than good.

Stealing Test Questions from Unsupervised Online Exams: Pre-Employment Screening Tests

In the area of industrial/organizational psychology, some pre-employment screening test items have been stolen as well, not so much because of the failure of security systems as in the case of IT certification exams, but because of a purposeful lack of security. For arguably strong business reasons, pre-employment screening exams are distributed widely over the Internet to potential new hires. A test may be taken by tens or even hundreds of thousands of candidates (see Burke, 2006). Scores from the tests are used to screen applicants: those with the higher scores make it to the next round in the selection process. The importance of making it to the next round of applications, especially during difficult economic times, is very high and people are highly motivated to achieve the highest score possible. Since the test questions are available without security on the Internet they can be copied from the browser and pasted into documents for easier distribution and use. The questions are drawn from a larger pool of items, thus requiring more retakes of the exam or better coordination between the thieves to capture as much of the pool as possible. Once captured, these questions can be shared with anyone who needs them to help answer the questions.

Several testing programs using pre-employment screening exams follow this same model. Although deliberately disclosing questions to an exam with such important consequences would seem self-defeating, there are several reasons why programs might do so (Burke, 2006). First, they need to reach a large number and wide variety of applicants. The use of the Internet to reach that audience is critical. Second, the test serves a screening function only and is not the basis for the final decision on who gets hired for a particular job; therefore it does not have the same weight as the use of a single test for hiring. Applicants who cheat in this first step may be discovered in a later step. Third, it is possible to use a proctored “verification” test at a later time in the process when the person is selected to be interviewed. That verification test covers the same content as the initial pre-screening exam but is given under more secure conditions. Despite the likely accurate assertion that the pre-screening test does not lead to incompetent people being hired, there are two facts that stand out. First, being without proctoring, questions are easily stolen and can be used by individuals to take the pre-screening exam. Burke (2009) reports that web patrolling efforts have revealed 18 sites over an 18-month period that contained actual items. Four of those sites were located in the UK, the other 14 in China. If applicants use these sites and take the exams, the results ultimately lead to screening errors with some people undeservedly advancing to the next step. More importantly, it often means that some individuals who should have advanced, but who did not
cheat, are unable to do so. There are several issues here. First, there are concerns about the fairness of the process and the ethical issues of psychologists creating and administering a purposefully non-secure, but important, exam (see Foster, 2010). Second, the process motivates, perhaps even encourages, some non-cheaters to accept the situation and begin cheating themselves in order to make the cut. Third, with an unknown number of individuals using stolen questions, how viable is the pre-screening process anyway? There may be many more cheaters making it through the screening process than there are positions available. Without proper security, the original purpose of the process may have been subverted.

Memorizing and Sharing Questions: College Admissions Tests and Licensure Tests

Memorizing test questions and then recalling and recording them later is a common way to steal test questions. As was briefly described earlier in the chapter, nearly 20 years ago Kaplan sent 20 of its employees to take the new computerized adaptive version of the Graduate Record Exam (GRE), owned by ETS, with the purpose of memorizing the questions and reproducing them later. Kaplan claimed to be evaluating the security of the new exam and that it never intended to use the questions as part of its test preparation business. A copy of 150 questions was sent by Kaplan to ETS to prove that the new computerized format made it easy to memorize questions which could be shared with other students. There was no report on how accurate the questions were relative to the actual counterparts on the GRE; however, ETS filed a copyright infringement lawsuit against Kaplan which was eventually settled for $150,000.

In 2002, the GRE was the object of another breach of security for ETS, which discovered websites with copies of questions and answers from the current operational version of the GRE. ETS stopped the use of the computerized GRE in several Asian countries, reverting to the paper-and-pencil version in those locations instead. It was unreported, and perhaps unknown, how the questions were stolen.

Other educational admissions exams administered by computer have had items stolen as well (Hechinger, 2008). In 2008 the Graduate Management Admissions Council (GMAC) cancelled the scores of 84 students because of evidence the students supplied or had access to stolen test questions for the Graduate Management Admissions Test (GMAT), posted on a Chinese-based website, ScoreTop.com. By seizing the ScoreTop.com hard drive, GMAC had “airtight” cases against the 84 students. At least 12 of the students had posted questions on the website, while at least 72 had accessed the questions. Again, the ScoreTop website guaranteed passing, so it was reasonable for candidates to realize that the test preparation materials were illegally obtained.

In a published response to inquiries (Lavelle, 2009), GMAC actually provided the text of one of the questions posted on the website as well as the actual test question from the GMAT. It is an important piece of information showing the accuracy between the actual test question and the memorized version of the question. Commentary by the test question thief in the posted item is italicized and shown in Figure 3.2.

Comparing the two versions, it is clear that they are not identical. Not all of the options were able to be recalled. None of the articles referring to this breach indicated the number of items that were disclosed, and except for the example item provided,
there was no indication of the overall accuracy of the posted items. However, enough information was posted to help a student prepare for the question in the event it was presented during the GMAT.

Health-related licensing programs have also fallen victim to item theft (Smydo, 2003). In 2002, the Federation of State Boards of Physical Therapy (FSBPT) discovered that students were sharing more than 100 questions on an online chat site. The students were recalling questions that they remembered from the exam, or asking other students to recall them. The FSBPT immediately took action to remove the offending material. Similarly, the National Boards of Pharmacy announced in 2002 that it had filed lawsuits against 15 “foreign-trained” pharmacists who posted and discussed more than 200 questions on two websites, one Korean and one Indian. Finally, the National Board of Podiatric Medical Examiners (NBPME) refused to validate the scores of hundreds of candidates because of concerns they may have had access to test questions. Keith Harris, the attorney for the Chauncey Group, International, NBPME’s testing vendor at the time, stated that “some students traded exam material by e-mail while others had access to a study guide containing questions ‘remarkably similar’ to those on the licensing test” (Smydo, 2003).

The increased number of test thefts of technology-based exams has concerned other disciplines as well. Psychologists wondered whether such trends would affect their discipline. Concerns were reported by the Monitor on Psychology, published by the American Psychological Association:

Figure 3.2 Compromised question posted on ScoreTop.com website (reprinted with permission from Caveon Test Security)
There is no reason to think that it’s affecting one discipline and not another,” says Leon Gross, PhD, director of psychometrics and research and associate executive director for the National Board of Examiners in Optometry. In fact, the growth of both professional testing and computer applications has only increased opportunities for testing violations, say experts.

“We are aware that there is more and more item exposure than ever before,” says Randy Reaves, JD, executive officer and general counsel for the Association of State and Provincial Psychology Boards (ASPPB), the organization that administers the psychology licensing exam.

The power of the Internet makes it easy for test-takers to share information quickly with countless other test-takers. In addition to videotaping exam rooms during tests, ASPPB also keeps tabs on Internet exam resources. “There is a chat room devoted to the exam, and we monitor that on a regular basis,” says Reaves. Obviously, offers of test questions or answers raise a red flag, he notes.

“Because people are taking the test by computer, that increases the opportunity to memorize items,” Reaves adds. Currently, ASPPB is investigating a number of very low scores – lower than even a “chance” score. If the scores are all in one jurisdiction, or there are other signs of possible cheating, the group will take action (Holloway, 2004: 28).

Other Methods of Stealing Test Content: University Exams

It is clear that memorization of test content is a relatively easy and undetectable way to steal test questions, and appears to be the method of choice when there is no easy way to capture the questions directly. Incidents of test theft of paper-and-pencil tests indicate that other technologies may be commonly used to capture test questions. These include those listed in Table 3.1.

There are many instances of documented methods of stealing test content from technology-based tests. Tests can be stolen by using new technology devices, or by tried-and-true methods of writing down the questions on scratch paper or by memorizing them. Some of these methods produce more accurate versions than others. For example, using a digital camera will produce an accurate version of a test question, without the answer.

It is not always obvious for breaches of technology-based tests to determine which method was used to originally steal the questions. Did the thief use a cell phone camera or a digital camera? Did they write the question down on scratch paper, or speak the text of the question quietly into a digital voice recorder? However, to the extent that a program determines how theft occurred, these programs need to be more forthcoming regarding the methods the thieves used to steal questions. Research efforts need to establish the effectiveness and popularity of these methods and what we can expect in the future.

In an interesting experiment to try out new technology, Sorenson (2008) obtained permission to try to “steal” questions from an exam given by a colleague at Utah Valley University in Orem, Utah. Armed with two new off-the-shelf recording technologies, a camera hidden in a button and a DocuPen, he attended the testing session, which was a paper-and-pencil test. Figure 3.3 shows both the ButtonCam and storage unit on a table. They sold at the time for about $400. Today, just a few years later, the
cost is less than $100 and the quality of the cameras is much higher. Figure 3.4 shows the button cam hidden as a button on the shirt. As the reader can see, the camera is undetectable by visual inspection. Figure 3.5 shows the DocuPen, about $250 at the time (less than $100 today), in action scanning a paper-and-pencil test form. From Figure 3.5 it can be seen that the DocuPen is larger than a normal pen and perhaps should have been viewed with suspicion by a proctor. However, the proctor did not question either device and allowed Sorensen to take the test.

The DocuPen works by swiping the pen across the page (or screen if this had been a computerized test) while in record mode. Sorensen did this several times without the detection by in-room proctors.

The images produced by the ButtonCam were not clear enough to read the text, making its use of questionable value for stealing questions, at least on a paper-and-pencil test. Perhaps it would have been more capable of recording the screen content of a technology-based test. However, the technology for digital photography is advancing rapidly and limitations of resolution are no longer an issue. For this research the DocuPen produced very clear images of the test, as is shown in Figure 3.6. It was very easy to capture the test pages in digital form without significant threat of detection. It is expected that these devices would have been similarly effective had the test been administered on a high-resolution monitor rather than paper and pencil. Colton (1997) provided a good review of capture technologies as they existed more than a decade ago. Today, more capable mobile devices such as smartphones and tablets not only provide unlimited Internet resources, but also have the ability to capture questions directly and quickly send them to others via instant messaging or emailing. Digital cameras continue to reduce in size and increase in resolution, making it easier than ever to capture test questions on an ongoing exam without fear of detection.
Figure 3.4 ButtonCam and DocuPen as they appeared when Sorensen entered the examination room (reprinted with permission from Caveon Test Security)

Figure 3.5 Using the DocuPen to record test questions (reprinted with permission from Caveon Test Security)

Figure 3.6 Images of paper-and-pencil test produced by the DocuPen (reprinted with permission from Caveon Test Security)

23. People who hold the __________ paradigm think only of getting what they want, are only set on winning, and don’t care whether you win or lose.
A. Win-Lose
B. Win-Win or No Deal
C. Win
D. Lose-Win

24. Which statement best describes the essence of Habit 7?
A. Effective, long-term relationships require mutual benefit.
B. Continuous renewal helps activate the other 6 habits.
C. We consciously choose our responses based on our values.
D. Quadrant III holds the essence of Habit 7.

15. What is the underlying principle of Habit 5?
A. The whole is greater than the sum of its parts.
B. Mental creation precedes physical creation.
C. Diagnosis must precede prescription.
D. Effectiveness requires balancing important relationship, roles, and activities.
Cheating

There are many different ways to cheat, as described in Table 3.2. The categories are ranked from most serious to least serious based loosely on the factors of opportunity, frequency, and effectiveness.

Obtaining Test Content Prior to Exam: Educational Admissions Testing, IT Certifications and Citizenship Exams

Some tests for college admission at the graduate level are administered by computers. Prominent examples include the Graduate Record Exam (GRE), the Graduate Management Admissions Test (GMAT), and the Medical College Admissions Test (MCAT). One of the most recent breaches (Hechinger, 2008) concerned the GMAT, described in the section above as an example of theft. The Graduate Management Admissions Council (GMAC), as a result of a legal ruling, was able to seize the hard drive for the website, ScoreTop.com. The hard drive contained email messages for 5,000 to 6,000 high-level subscribers to the website with access to the ScoreTop versions of the test questions. After reviewing the data on the hard drive, GMAC concluded that 72 GMAT test takers had admitted to viewing versions of some of the actual GMAT questions that were posted on the site. GMAC cancelled the applicants’ test scores and allowed them to retake the exam. It is not known how many others of the several thousand subscribers had posted questions or seen the posted questions. Forensic analyses of the GMAT test results that may shed more light on the extent of cheating, if conducted, are not publically available.

The most effective cheating occurs when a person is able to obtain exact copies of the test questions and answers prior to the test. To go into the test with all or some of the questions and answers is a major advantage over other test takers and virtually guarantees that a test score will be enhanced significantly. Often it only takes an advantage of a few questions to result in a passing score or to make the difference in a grade or the chance of admissions into the college of choice.

The previous section discussed how items can be stolen and in some cases how those items are made available to test takers. One common method of distribution of stolen questions is the use of braindumps, websites that sell what they call “test preparation” materials. As research has shown (Foster & Zervos, 2006) these materials are, for the most part, identical copies of test questions.

Two important questions that must be addressed are how commonly used are braindump sites and whether they are effective. In one of the first methods for detecting and calculating the effects of question pre-knowledge, Maynes (2009) provides research on actual cheating rates and their effects. He reported the analysis of test results from three major IT exams for three different IT certification programs with a history of battling test theft. Working with Maynes, the three programs had seeded easy, previously undisclosed test questions, usually fewer than ten, into an existing exam. The new items, termed Trojan Horse items, were deliberately miskeyed; however, those items were also unscored for purposes of making the certification pass/fail decision. When the test files were stolen, the key was stolen as well. The wrong keys would then be provided to users by the braindump sites. The logic behind the Trojan Horse approach is as follows. History with these programs suggests that, shortly after a test is launched, the test content will be stolen and published on a
### Table 3.2 Types of cheating

<table>
<thead>
<tr>
<th>Brief description</th>
<th>Reason for ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Obtaining test content prior to exam</strong></td>
<td>The content is highly accurate and easily obtained. The actual cheating is very difficult to detect.</td>
</tr>
<tr>
<td>Test taker obtains a copy of test questions and answers prior to the exam. These</td>
<td></td>
</tr>
<tr>
<td>may be obtained from websites selling the exact questions, or from forums and chat</td>
<td></td>
</tr>
<tr>
<td>sessions where some or many of the exams are discussed. Some questions may be</td>
<td></td>
</tr>
<tr>
<td>obtained from a friend who took the test earlier.</td>
<td></td>
</tr>
<tr>
<td><strong>2. Colluding with expert while taking the test</strong></td>
<td>This type of cheating is not detected because it often involves collusion with testing center personnel.</td>
</tr>
<tr>
<td>With the help of the testing center personnel, experts in the tested content assist</td>
<td>High scores are assured.</td>
</tr>
<tr>
<td>the test taker during the exam. These helpers may be actual teachers, instructors,</td>
<td></td>
</tr>
<tr>
<td>or simply knowledgeable coaches. It some cases it may involve communication with</td>
<td></td>
</tr>
<tr>
<td>the test taker via cell phone or two-way radio.</td>
<td></td>
</tr>
<tr>
<td><strong>3. Using inappropriate or non-authorized test aids</strong></td>
<td>Test taker is able to access resources to cheat during the test. The use of these resources is very</td>
</tr>
<tr>
<td>Test takers use non-authorized aids during the exam, such as cheat-sheets, cell</td>
<td>effective and they are designed to be difficult to detect. The resources may be helpful but not very</td>
</tr>
<tr>
<td>phones, headphones, programmable calculators, etc. Test takers may be able to access</td>
<td>accurate. A high score may not be assured.</td>
</tr>
<tr>
<td>the Internet and its resources during the exam. These devices and resources contain</td>
<td></td>
</tr>
<tr>
<td>test questions/answers or simply information about the subject matter being tested.</td>
<td></td>
</tr>
<tr>
<td><strong>4. Using a proxy test taking service</strong></td>
<td>This method of cheating has the assurance of a high score, but it is expensive. Professional proxy</td>
</tr>
<tr>
<td>It isn’t difficult to find a company offering proxy testing services. These expert</td>
<td>test taking is difficult to detect as often it involves collusion of testing center personnel or very</td>
</tr>
<tr>
<td>individuals will take the test for you. The services may even be located in another</td>
<td>realistic, but fake, identification documents.</td>
</tr>
<tr>
<td>country. There is no need to go to the test center or attend the test session at all.</td>
<td></td>
</tr>
<tr>
<td>These services often guarantee a high or passing score.</td>
<td></td>
</tr>
<tr>
<td><strong>5. Using friends or acquaintances for proxy test taking</strong></td>
<td>This type of cheating is difficult to detect when it takes advantage of lax ID-checking processes.</td>
</tr>
<tr>
<td>For a smaller price, or even as a favor, a friend or acquaintance may take the test</td>
<td>There is no assurance of a high score.</td>
</tr>
<tr>
<td>on behalf of the cheater.</td>
<td></td>
</tr>
<tr>
<td><strong>6. Hacking into a database to change a score</strong></td>
<td>It is a difficult process to access a program’s systems, requiring strong technical IT skills and</td>
</tr>
<tr>
<td>Exploiting weak user access systems, hackers can gain unauthorized access to a</td>
<td>knowledge. The probability of detection is moderate to high, with the threat of serious</td>
</tr>
<tr>
<td>program’s test results database or scoring system. Once in the system they are</td>
<td>consequences.</td>
</tr>
<tr>
<td>able to change lower scores to higher ones.</td>
<td></td>
</tr>
<tr>
<td>**7. Copying the answers provided by another test taker while both tests are being</td>
<td>This type of cheating is low risk but is mostly ineffective today with technology-based tests because</td>
</tr>
<tr>
<td>taken**</td>
<td>most current computerized tests are designed such that test takers do not see the same questions nor in</td>
</tr>
<tr>
<td>Even for a computerized test, it is possible to copy answers from a nearby test</td>
<td>the same order.</td>
</tr>
<tr>
<td>taker. Questions may not utilize randomization features, making this method of</td>
<td></td>
</tr>
<tr>
<td>cheating more effective.</td>
<td></td>
</tr>
</tbody>
</table>
braindump site. Purchasers of the items then memorized them, including the answers, and took the test. Logically, those using braindump sites will do very well on the scored items of the test, even passing the test, but will do very poorly on the Trojan Horse items. This is because, presumably, candidates relying on braindump sites have very little knowledge of the subject matter, and the wrong key was published for the Trojan Horse items. Using the Trojan Horse technique to detect cheaters with pre-knowledge, Maynes found that 35, 40, and 85 percent of the test results for the three exams showed strong evidence of the use of braindump files. These percentages are obviously very high. He also reported that 93 percent of the test takers – for the exam where 85 percent had cheated – passed the exam needed for certification. Clearly this method of cheating was very effective. For those who were identified as not using pre-knowledge, the pass rate was 53 percent. Although there is no way to know whether these examinees would have passed the exam had they not had prior access to the test questions, there is ample reason to suggest that many would not have. Do, Brummel, Chuah, & Drasgow (2006) provide additional support for both the real and perceived value of pre-knowledge. In an experiment with 386 undergraduate students, the researchers provided some students with pre-knowledge of items on a cognitive ability test and compared their test performance with test takers that had no pre-knowledge. The pre-knowledge group had previously been given a “cheat sheet” compiled from memorized questions recalled by previous test takers. Some of the pre-knowledge was accurate and some was not. They found a “moderate to large” positive effect of the pre-knowledge. That is, the “cheat sheet” was effective in helping the subjects achieve higher scores. In addition, they found that subjects relied heavily on the information on the cheat sheet, trusting that information much more than their own knowledge or ability.

In 2008 a language school in the UK was discovered providing pre-scripted answers to tests of proficiency in the English language and for tests of English culture (Kemp, 2008). The test is required of individuals seeking British citizenship. The course associated with that exam normally took 150 to 200 hours to complete. The language school described by Kemp provided the questions and answers in a single afternoon at a cost of £900. It provided answers and instructions on how to pronounce the words they would see. There were no data reported on how many individuals purchased the course, nor on how effective cheating was. But one can only assume that those with access to live test content ahead of time were greatly advantaged.

Unproctored Internet Testing: Industrial/Organizational Psychology

Since the methodology of purposeful unproctored testing has come up over the past few years, there are few studies evaluating unproctored Internet tests. In the area of pre-employment screening exams Do, Drasgow, & Shepherd (2007) compared unproctored Internet tests for cognitive ability tests at two different times, the logic being that over time test content would be disclosed leading to increased cheating and rising test scores from the first to the second instance. Other measurement properties of the test should change as well. Over a two-year period, however, only minor score inflation was discovered. While the results would suggest that unproctored online testing may not benefit from security measures, it was suggested by the authors that the stakes of the tests may not have been high enough to motivate cheating. In other
research, Davies & Wadlington (2006) compared two “nearly identical groups” of motivated job applicants on personality tests. One group had proctored exams and the others were unproctored. While the method and vigilance of the proctoring wasn’t described, they ultimately found no important differences in test score means or item performance. Some researchers have suggested that personality tests, having essentially no right or wrong answers, are somewhat protected from the efforts of cheating and test theft. Finally, in the area of distance learning, a study by Harmon & Lambrinos (2008) compared tests from two online courses. For one course the final exam was proctored; in the other it was not. From the results they concluded that higher scores were achieved and that cheating was taking place when exams were unproctored. To date, the studies in this area are few and the results contradictory.

While not comparing proctored to unproctored tests, Burke (2008) applied forensic indicators of test fraud to the data from the SHL unproctored pre-employment screening cognitive ability test. The individuals taking the SHL screening test understood that if they passed they might have to complete a proctored verification exam. It was suggested that this possibility (taking a verification exam) may deter individuals from cheating to pass the screening exam, although no data exist to verify this effect. Evaluating the unproctored exams, Burke applied multi-factor data forensics methodology to almost 30,000 test results. Table 3.3 shows his results. Overall, less than 2 percent of candidates were flagged against at least one data forensics index. Burke cautions that while 2 percent in total may seem small, it still translates to many applicants for large testing programs. It is important to mention that these forensics methods have unknown power; therefore it was likely that there was some cheating that remained undetected. If only a small number of individuals are needed to fill job openings, it is uncertain what percentage of those might be filled by the cheaters.

Another unproctored online pre-employment screening test has come under criticism. Kronos, Inc., administers an exam that screens hourly retail workers, one of the largest classifications of workers. The test – the Frontline Reliability Assessment (also called the Unicru test) – predicts “whether a job candidate has an elevated likelihood of being terminated for delinquent behavior and is likely to leave after just a short time on the job” (Business Wire, 2005). The company regards the test as secure and believes that it can be given online and unproctored, primarily since the test is a measure of personality traits and therefore the correct answers are not easily discerned; however, it provides no data to support the assertion that few or no individuals cheat. The Frontline Reliability Test is objectively scored,

<table>
<thead>
<tr>
<th>Data forensics index</th>
<th>Index value (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collusion</td>
<td>0.21</td>
</tr>
<tr>
<td>Identical tests</td>
<td>1.12</td>
</tr>
<tr>
<td>Perfect score</td>
<td>0.7</td>
</tr>
<tr>
<td>Response aberrance</td>
<td>0.09</td>
</tr>
<tr>
<td>Score aberrance</td>
<td>0.03</td>
</tr>
<tr>
<td>Fast latencies/high accuracy</td>
<td>0.003</td>
</tr>
</tbody>
</table>
with results being provided immediately. Therefore, correct answers are inherent in the scoring and available to be determined in several ways. A quick search on the Internet for answers or strategies for the exam lists a large number of sites with answers, copies of questions, and strategies for answering the test. Even without knowing the answers, an applicant can simply have someone else take the test for him or her. Without authentication and supervision there is no way to make sure the applicant is actually the person taking the test and that cheating is not happening. Retakes are encouraged, as indicated in an article in the *Wall Street Journal* which provides an analysis of the security problems for this exam (O’Connell, 2009). One of the links in the article leads to a website by Horrigan (2009) offering very specific advice on taking the test. To illustrate the strategies, as well as the ease of cheating and the brazenness of applicants cheating or helping others, much of that text is provided here:

Many of these stores use a seemingly inscrutable personality test called the “Unicru” test. Even though I tend to be good at standardized tests and have even worked in the testing industry for a while, I have flunked this one a time or two. Happily, I have found one of the many answer keys which is floating around the Internet.

This is a totally unofficial key, compiled off the top of my head from my limited experience and study, using random materials which have been floating around the web for years. Use it at your own risk. I got the basic info from a blogger … I tweaked it by sorting the prompts by the expected answers – which are always “Strongly Agree” or “Strongly Disagree.” … “Disagree” or “Agree” are NEVER the right answer to any question, even though any sensible person will have mixed feelings about all these questions.

There are 99 questions in this list. I have taken this test several times, and I am pretty confident all these questions are still in use, and I don’t specifically recall any other questions. You will not have to answer all 99 of them, but you will be asked about half of these questions. If you get enough right answers, you go in the Green pile and you may be called in for an interview. If you get almost enough right answers, there is a Yellow pile, which is used if no one from the Green pile wants the job …

The general drift of the test is to try to make sure that you are hard-working and honest but not too ambitious – and you must be cheerful all the time and enjoy being around people all day. The answers are mostly straightforward.

Horrigan goes on to provide text for nearly 100 questions and how to answer them. It is unknown how many individuals cheat on the Frontline Reliability Assessment/Unicru. If the number of websites providing test questions and strategies for cheating is any indication, the number of cheaters may be quite large. In light of this, it is very questionable as to whether the results can be used to screen workers fairly and competently.

*Using Inappropriate Test Aids: Licensing Exams*

Williams (2009) reported that “outside materials,” consisting of cheat sheets or note cards, were used by prospective emergency medical technicians to help answer questions on the exam provided by the National Registry of Emergency Medical
Technicians taken on computers at a Pearson/VUE testing center in Maryland. Chavez (2009) added additional information regarding the testing center. He indicated that the popularity of the Maryland site was partly due to the fact that the security at the center is not strong. One firefighter cadet described the site: “It’s more relaxed, period. Nobody over your shoulders.” The site is a preferred site for the exam compared to sites closer to where the candidates are located. Sources vary with respect to estimating the number of cheaters; one source indicates between 45 and 60, while another says as many as 70 to 200 firefighters were involved.

Proxy Test Taking: IT Certifications, College Admissions Testing, Tests of English Competency

Recently for large sums of money, training faculty and system administrators colluded in impersonating over 100 certification candidates in India, taking the tests for them and helping them to obtain the Cisco Certified Network Administrator certification or CCNA (Pradesh, 2009). Even the director of the testing center was aware of the proxy testing service. The specific procedure required that the candidate appear at the testing center at the scheduled time and day. According to the report, the candidate would have a photograph taken with a webcam and sign a digital signature pad. At that point, a center administrator impersonated the candidate. Since the proxy test takers were also instructors in the technology, it was easy for them to answer the questions correctly and achieve a passing score on behalf of the candidate.

It is not always necessary for the candidate to actually appear at the testing location. Proxy test taking is advertised worldwide on the web to be purchased by anyone. One such website is blatantly called www.BuyITCert.com. It is based out of Hong Kong and promises (LIVE-PR, 2009) that the customer can receive “the IT certification without appearing for any test or attending the training.” The website continues, “Passing the exams is guaranteed or else they will refund you the complete fees collected for the exams. Amazing, isn’t it!!” A set of five exams for a Microsoft security certification is priced at $3,200 and can be paid in four installments. BuyITCert advertises that the exams are taken at authorized testing centers for Prometric and Pearson/VUE.

Caveon Test Security evaluated the validity of the claims of one of the many websites similar to www.BuyITCert.com. Maynes (2009) reports that Caveon was contracted by an IT certification program to purchase the services of a proxy testing organization. Two weeks after sending the payment, Caveon was informed that the test was taken and passed, and the certification awarded. Caveon subsequently learned through its data forensics analyses that specifically look for such testing that a single proxy testing service, over six months, took tests for over 500 candidates at six different “authorized” testing centers. At $1,000 per exam taken, this adds up to revenue of $1,000,000 over the course of a year – for a single exam title!

It is not only the IT certification industry that is plagued by such services. In 2003 there were nearly 600 cases of proxy test taking of the GMAT (Lavelle, 2009) for $5,000 each. Proxy test taking has been an ongoing problem for international tests of English competency, such as the Test of English as a Foreign Language (TOEFL) and International English Language Testing System (IELTS; Jopson and Burke, 2005). In 2009, in India, a rise in available impersonators was reported for the TOEFL (The Hindu, 2009).
Using Friends or Acquaintances for Proxy Test Taking: Military Entrance, Employment Screening, and Distance Education Course Exams

The American Armed Services Vocational Aptitude Battery (ASVAB) is the main instrument the US Department of Defense uses to select and classify recruits. As a computerized test, it is administered at more than 65 Military Entrance Processing Stations (MEPS). In 2008, a reporter for the Houston Chronicle (Schiller, 2008) described a cheating scandal where several military recruiters located and used “stand-ins” to take the tests for recruits who would not otherwise have qualified for the US military. The specifics of the scheme have not been released by the military and it is not known how many people have entered the military in this way. Information suggests that proxy test takers were used for at least 15 persons at the MEPS in Houston. The recruiters worked at a total of four MEPS. One of the recruiters caught in the scandal, who had put a total of about 65 people into military service, had used this method six times. The recruiters, who either left the service or were discharged, maintained that they acted at the request of and under the approval of higher ranking individuals, that this type of practice has been going on a long time, and is often necessary when there is pressure to increase the number of people entering the military. Part of the justification reported by one recruiter was that the individual would still have to complete advanced training school and bootcamp.

Exams with important consequences that are unproctored or where there is lax proctoring are susceptible to allowing a friend or acquaintance take the test for the actual test taker. This type of behavior occurs often in college testing settings where an instructor may be the only person monitoring a large number of students. Proctors in many testing center networks are usually part-time, underpaid, unmotivated, or have a vested interest in how the test taker performs (Cizek, 1999). These factors combine to make it easy for a friend or acquaintance to take the exam. Identification procedures often are not followed properly or carefully, or are not used at all. For unproctored exams given over the Internet, the ability to take tests for another person is even easier. By the very nature of being unproctored, these exams have no way to authenticate the identity of the test taker. Many, perhaps even most, of these exams may be taken by a person other than who is supposed to be taking the exam.

Without strong authentication processes, it is hard to understand how the results of the test can be considered valid and used for hiring decisions.

Recently the US government passed the Higher Education Opportunity Act, also called the College Opportunity and Affordability Act (HOEA, 2008). The Act recognizes that, in the fast growing field of distance education, mainly education provided online, there is a need to authenticate that the person obtaining a grade in a course or a degree is the person who actually completed the work, including completing exams. Under HOEA (2008), beginning in 2010, institutions offering distance education are required to have “processes through which the institution establishes that the student who registers in a distance education course or program is the same student who participates in and completes the program and receives the academic credit.”

College athletics is a prime breeding ground for proxy test taking. In a well-publicized recent example of cheating, 39 Florida State University student athletes admitted to receiving help in 2006 and 2007 during online exams for a music course (Dinich, 2009). The helper, described as a “learning specialist, academic advisor and
tutor,” provided answers to online course quizzes and by asking another player to complete it for the athlete registered for the course.

**Changing Test Scores: Military Entrance Testing and Educational Admissions**

In 2007, in another case involving the ASVAB, a test administrator and proctor took bribes and changed the test scores for about 70 Army National Guard recruits. Tan (2007) reports that Christine Thomas, a federal employee, began taking bribes after being on the job for six months. Some of the requests to manipulate the scores came from recruiters and not from the recruits themselves. She charged an average of $21 and made about $1,500 over 18 months. It is important to note that a spokesperson for the Arizona National Guard, downplaying the seriousness of the fraud, remarked that ASVAB scores are not as important as the schooling that a recruit receives.

In March 2008, Harvard University officials announced that a database containing student information had been hacked (Vamosi, 2008). Of most concern were 6,600 summaries from United States admissions candidates which included the applicant’s name, Social Security number, date of birth, address, e-mail address, phone numbers, test scores, and other academic information. The data were placed on a publicly available BitTorrent site. Admissions test results are kept in several different databases, including the admissions testing program, the test administration vendor, and the school to which the scores are sent. With the increased sophistication of hackers, and the difficulty in keeping up with the threats, it is not hard to imagine that many of these databases are less secure than they should be.

**Solutions for Test Fraud on Technology-Based Tests**

**Prevention and Deterrence**

Prevention and deterrence are two strategies for reducing security problems. In the testing context, prevention is a solution and removes the basic cause of the security problem completely, making it impossible or very difficult for the problem to occur. Benjamin Franklin’s well known saying that “an ounce of prevention is worth a pound of cure” also applies in testing situations. If preventing cheating from occurring is possible, and if prevention can be implemented at a reasonable cost and without privacy or legal restrictions, then it is a preferred way to deal with specific test fraud.

Deterrence methods, on the other hand, are ways to psychologically prevent test security fraud by providing the test taker and/or others with the perception that they will be easily caught and dealt with severely if they try to cheat, steal test content, or share/sell such content to others, or that what they have stolen will not be useful to themselves or others. According to Wells (2002) in an article published on the Certified Fraud Examiners website, deterrence is the “modification of behavior through the threat of sanctions.” He goes on to explain that the perception of an increased likelihood of getting caught is more likely to deter inappropriate behavior than the reality of having in place detection mechanisms. It highlights the importance of making sure that all those involved in the testing process are informed about and realize that highly effective detection methods are in place, that stolen content will have limited usefulness, and that aversive consequences will be swiftly and fairly applied when a person is caught.
Specific Prevention and Deterrence Solutions

Testing programs, both paper-based and technology-based, rely on both prevention and deterrence methods. The delivery mechanism in TBT affects both the vulnerabilities and the strategies for addressing those vulnerabilities. Some examples of how prevention can be effective with TBT follow.

Test File Theft

This section provides solutions for test files downloaded to a testing center, stolen in various ways from unproctored tests on the Internet, or obtained from a testing program insider. It is important also to adhere to more general security solutions that apply to both paper-and-pencil and technology-based tests. These are provided in other chapters in this Handbook.

Protect Test Files

When entire test files and supporting graphics, video, or audio files need to be downloaded to a testing center, strong encryption processes will prevent thieves from accessing the test information, even if they are able to access and capture the files. One common system is symmetric-key encryption. Symmetric-key encryption provides computers with a secret key or code that is used to encrypt information before it is sent over a network to another computer. If the receiving computer also has the key or code, it can decode the message. 128-bit encryption, in common use today, means that the key used to encrypt the message is 128 bits long and has $3.035 	imes 10^{38}$ possible combinations. That length is sufficient to prevent random or casual guessing of the code; however, any code can be broken if enough computing power is applied to the task. As these security encryption methods are being improved continuously, usually in response to new threats, the program is advised to remain informed and up-to-date on what is available, or to take advantage of the experience of certified IT security consultants.

Download Only Required Items

Often, the entire item pool supporting multiple test forms is sent to a test center in preparation for a scheduled exam, even though only one test form is scheduled to be provided to the test taker. To prevent unnecessary capture of questions and other test content it is recommended that the test publication and testing system only download the necessary items to support a particular test form. Of course, if the test design requires selection of items from a larger pool (e.g., CATs or variations of them) then the entire pool will need to be downloaded and available for use.

Just-In-Time File Transfer

If test files must be downloaded to a testing center prior to the exam, they should remain on the server for an absolute minimal amount of time. Prior to the test, the files should be downloaded to the local server only a short time before the test is scheduled. This decision may depend on particular transmission conditions, especially
in certain parts of the world, such as the availability and speed of the local Internet connection. Similarly, when the test has completed it is critical that the test files be removed completely. This is not as simple a task as it sounds because files, even if deleted, can be easily recovered. Dishonest testing center personnel may not delete the file as required or may move it to a new location. It is important for programs to verify the files have not been copied, moved, or renamed, and that they have been permanently deleted. To make sure the file is actually deleted, any one of a number of tools can be used. Eraser (2012), for example, is easy to use and free. It overwrites with random data patterns until the file is no longer recoverable.

Create Larger or Unlimited Item Pools

Particularly after the Kaplan/GRE incident, testing programs have been advised to increase the size of item pools (Parshall, 2002; Davey and Nering, 2002) or to increase the number of questions in their tests or in the question pools supporting their tests. For this procedure to be effective it would be very important to be able to create useful items as quickly and inexpensively as possible, and to continue to refresh the test forms with new questions. Some content areas (such as mathematics) lend themselves to automated item generation, even on the fly during an exam. Bejar’s (1996) “generative response model” uses an extensive analysis of a content domain and psychological processes to generate new questions and associated statistical parameters, bypassing the traditional approach of pretesting questions and analyzing their data statistically. Van der Linden and his colleagues (van der Linden, 2009; Hartig, Harsch, & Höler, 2009; Geerlings & van der Linden, 2009) present the theory and logic behind “item families” and provide research supporting a model that would automatically create large numbers of test questions.

For some testing programs, questions are compromised so quickly, sometimes within hours or days of the test’s publication and initial distribution, that they need to continuously produce new items to replace them. In addition to the automated item generation models, item cloning models may provide a reasonable way to quickly increase the size of an item bank (Hambleton, 2002). Cloning is a methodology that changes enough features of the item so that it is relatively unrecognizable to a cheater with pre-knowledge, but retains its statistical properties.

Finally, some have suggested that larger item pools prevent users from capturing (e.g., though digital photography) or memorizing enough of the item pool to have enough impact to compromise or affect the test’s reliability or inferences from validity evidence (Way, Steffen, & Anderson, 2002).

Control User Access

One way to prevent item theft is not only to control access to the item pools and tests during development, but also throughout the test administration process. When files are residing on a local server prior to or just after the scheduled test event, they should be protected from unauthorized access, in addition to being encrypted. At no time should anyone but an authorized and trusted test administrator have access to the files. New and stronger user access systems are constantly being developed and implemented, adding advances in features such as the use of biometrics. It may not be
unusual in the future to subject test administrators and proctors to the same security protocols as test takers. For example, to access the testing system a test administrator may be monitored in real-time through webcams and required to provide a biometric to authenticate him or her. Programs may employ the services of an IT security specialist in order to evaluate if their own systems, or the systems of their contractors/vendors, are secure from unintended access.

Use One-Item-at-a-Time Test Administration

It may be preferable to avoid the security problems associated with downloading test files by using one of the many Internet testing engines. While some Internet engines adhere to the download model, most do not, preferring instead to simply administer each item via a live Internet connection. With this model, only one item at a time is presented to the test taker, and no item is ever stored, even briefly, on the local hard drive or local server hard drive. The actual item files or pool of items are stored on a distant web server, generally unavailable to the thief. The items administered in this way are encrypted during transfer and decrypted just prior to presentation.

Internet Site Monitoring

Items are stolen routinely and shared with others. The Internet provides the likely vehicle for sharing those questions, either in chat rooms, forums, eBay, and dozens of other categories of sites. It is important for a testing program that uses technology-based tests to constantly canvass the probable locations on the Internet looking for the specific display or discussion of test questions. When such sites are found, a testing program must take immediate action to contact the owner of the site, let them know what is happening (they often are not aware), and ask them to respect the program’s intellectual property and take down the information. A surprising percentage will comply with this request, having been unaware that what they have been doing is illegal. However, there are some that will not agree because they know what they are doing, are making money, and have perfected ways to resist the pressure to take down the information. The testing program may have to take legal action quickly in order to remove the information, take down the site, and send a message to others who might try it. In a story with a happy ending, GMAC, the publishers of the GMAT, recently won a court battle and gained control of the website, ScoreTop.com, its hard drive containing the information of business school applicants who disclosed questions or had access to such questions, and won a judgment worth $2.35 million.

Rapid Repair and Republication

One luxury of technology-based tests is that if one or more items have been compromised, the exam can actually be repaired, republished, and distributed across the world, within a very short time period and relatively inexpensively. Although Internet technologies allow the realization of this advantage, many programs continue to use outdated software and services that make this solution difficult, impossible, or expensive. Often the process takes several weeks and costs thousands of dollars, inhibiting organizations from making these very necessary changes. This difficulty
increases the security risks and potential legal liability. A testing program should not have to tolerate a situation where they must continue to give an exam where the questions have been stolen and exposed on the Internet. Nor should they have to suspend the testing completely for several weeks, damaging their program and delaying the progress of examinees. Integrated or closely aligned item banking and test administration systems, with proper security procedures in place, allow for (1) detection that questions are not functioning statistically within acceptable bounds or have confirmed to have been stolen, (2) replacing the items in a test form with suitable replacement items available in the item bank, (3) recreating the test form with the new items, (4) republishing the new test form and making it available to all test takers, even those who had previously scheduled the earlier version of the exam. The technology needed to implement steps 1 to 4 quickly and at minimal cost widely exists, and programs should either utilize software that allows for this important security feature or work with their delivery vendors to begin to enhance their current systems to enable it to perform this activity.

Digital Watermarking

I know of no current operational effort to digitally watermark technology-based exams, even though the approach has merit. The ideal use of digital watermarking would be to code each instance of the administration of a selected subset of exam questions in such a way as to link it to an examinee taking the test. If the coded items appear on the web they can be easily traced back to the examinee, day/time, or location where the test was taken. The watermarking process should be systematic in using non-obvious, trivial, and relatively undetectable changes in the content of the items, such as option order (i.e., non-random), font sizes or spacing, minor wording differences, small changes to non-essential parts of a graphic, etc. These changes would occur on the fly, dictated by the characteristics of the individual testing instance. As a detection methodology, in conjunction with web monitoring activities, it can be a valuable tool in detecting and dealing with theft of entire tests. It would certainly be less effective for other methods of theft, such as when test takers memorize questions.

Test Item Theft during the Examination

Whereas the previous section focused on strategies for preventing or deterring the theft of test files or item bank files, this section focuses on preventing or deterring the theft of items during the exam itself. Again, the list presented here focuses on those security features that are unique to TBT, although the reader is reminded that it may also be necessary to implement many of the general security strategies that are applicable to all types of testing.

Control the Browser and Operating System

With any type of technology-based exam, it is very important that test takers only have the ability to answer the test questions and perform some test navigation functions during the exam, but that all other behaviors are prohibited or prevented. This is true for operating systems as well as Internet browsers. All unauthorized
avenues of escape from and re-entry to the test must be prevented. Browser and operating systems lockdown procedures make sure that a test taker cannot access other computer resources such as files on the hard drive or another website. Typical lockdown features include:

- There is no right-click menu.
- There are no browser control buttons (Back, Next, etc.).
- There is no access to printing resources.
- Prnt Scrn key is not active.
- Taskbar and desktop are hidden.
- There are no menu or program icons.
- There is no ability to minimize or maximize windows.
- Copy and paste text are not possible (except in writing assessments).
- Only a single instance of a test can run at a time.
- There is no ability to launch or access other applications.
- Use of certain keys (Ctrl, Alt, Fn, etc.) is prevented and logged.
- Test exit occurs at test completion, when the time runs out, or at test taker’s decision to end early.

Use Protective Item Design Features

Items can be designed to protect against test takers attempting to steal their content. This has generally been a difficult chore because the whole point of presenting a test question is to provide enough information for the test taker to competently respond if possible. With technology-based tests, it is important to review the capability of items. Even the standard multiple-choice item design can be reconsidered in light of the capability of new technology and security concerns.

Discrete Option Multiple Choice (DOMC) Item. A new item type (Foster & Miller, 2009) presents the standard content of a multiple-choice test question in a way that exposes less of the item to test takers. With DOMC, the options are randomized and presented one at a time, with Yes and No buttons for the examinees to indicate whether they believe the presented option is or is not the correct answer. An example item is shown in Figure 3.7.

The top panel of Figure 3.7 shows a traditional multiple-choice question with a stem and all of the options presented. In this case, only one option is correct. The bottom panel shows the same item presented in DOMC format. While only one option is shown at a time to the examinee, all of the options remain available to present if needed. The test taker’s task here is very simple: indicate if the option displayed is the correct one or not, by selecting the Yes or No buttons. When the examinee indicates “Yes” to an alternative (regardless of whether the option really is the correct answer) or indicates “No” to the correct choice, no more alternatives for the item are shown and the examinee moves on to the next question. The examinee does not receive feedback on whether the item was answered correctly.

In most cases, not all of the options are presented, which reduces exposure of critical components of the item while maintaining or even improving the measurement of the examinee’s knowledge or ability. The primary measurement advantage is that
examinees are unable to answer through process of elimination or evaluation of the available options, including less construct-irrelevant variance. The security benefit is that the non-exposure of multiple-choice options prevents their memorization and other forms of capture.

Research (Foster & Miller, 2009; Kingston, Foster, Miller, & Tiemann, 2010, in press) has recently demonstrated the security and other advantages of the DOMC item type. Scores obtained from tests made up entirely of DOMC items are generally several percentage points lower than the same test with traditional multiple-choice items. By reducing the usual construct-irrelevant advantages of cheating, scores will be lower, but will be more accurate representations of the test takers’ knowledge or ability levels.

Substitutions of Irrelevant Text. With technology-based tests, it is possible to substitute elements of text for a particular item each time it is presented. The text should be irrelevant to the content of the question, for example, the name of an individual described in a scenario. Such a change may reduce the usefulness of an item once it is stolen, as it will make it more difficult for the test taker with pre-knowledge about the item to recognize it.

Randomization of Options. Randomization, as a security tool, can be applied to the presentation of items or, for multiple-choice questions, to the presentation of options. Both applications deter the theft of questions and answers (i.e., copying) from another. A thief who steals just the item numbers and ordinal positions of answers cannot guarantee that when questions appear on subsequent exams they will appear in the same ordinal position, and if they are multiple choice questions, that the correct answer label (i.e., A, B, etc.) will be the same. Referring specifically to multiple choice options, Impara & Foster (2006: 102) suggest that “when there is no logical or numerical
order, then response choices should be arranged randomly. Moreover, when multiple versions of a test are used in the same testing window (or when computerized testing is being done), the response choices can be randomized within and across test versions.” They go on to state that such a strategy will have little effect on item difficulty or reliability, but will have a strong positive effect on security.

Use Protective Test Designs

Protective test designs are those that do not disclose all of the test questions to the test taker, making it difficult for the cheater to anticipate the questions he or she will view during the exam. In addition to that, though, such designs actually reduce the overall rate of question presentation, all other things being equal. There are several ways to do this.

Computerized Adaptive Tests. While CAT introduces some security vulnerabilities not inherent in other types of tests, as demonstrated by the ETS/Kaplan breach experience with the GRE, CAT does possess some security advantages such as reducing the overall exposure rates of items. It is not uncommon for adaptive tests to be only half as long as paper-and-pencil tests, while providing the same amount of information about a person’s trait level.

Implement Logical Stopping Rules. Technology-based tests can update an examinee’s score throughout the exam, after each question has been answered. For criterion-referenced tests, it is likely that an examinee’s classification (e.g., pass/fail) will be determined prior to the scheduled completion of the test. Nevertheless in virtually all technology-based tests, questions continue to be administered, disclosing their content unnecessarily. Once a pass/fail scoring criterion has been met, the test can and should stop. There are few psychometric reasons to present additional questions, but every security reason not to. In addition, for norm-referenced tests, stopping rules should be based on either reaching the maximum number of items or the minimum test information threshold, whichever comes first. Again, presumably for the scheduling convenience of CBT test centers, most programs administer fixed-length CATs, meaning that they continue to administer items after the examinee’s classification is known or their score has been measured with sufficient precision.

Avoid Overly Long Tests. Any item not included on an exam is one that can be used on an equivalent form or can be held back to later replace a compromised question. Many tests present more questions than are absolutely necessary for validity or reliability purposes. The number of questions in an exam may have been chosen arbitrarily or because it has “always been that way.” Psychometricians today are trained to design tests according to detailed specifications including attaining reliability and validity targets. Enough items need to be included to satisfy those requirements, as well as content specifications. After those needs are met, no more items should be included in the exam.

Use Multiple Equivalent Versions of Tests. Test designs that require different versions, or forms, of a test are effective at discouraging test theft. The logic is that stealing one
form of an exam will not be helpful if shared with others who will likely see a different form when they test. The forms utilize the total number of questions available for an exam, and do not reduce overall item exposure; they simply spread the items around in different ways. Two popular designs are Linear-on-the-Fly (LOFT) and Multiple Equivalent Forms. In the LOFT design, the items are selected from a pool for each test taker in a modified random fashion, according to specifications. For example, questions may be selected randomly from a set of questions in a particular content area that is needed. Programs using Multiple Equivalent Forms create as many forms as desired from the pool in advance of test publication. The forms are created according to psychometric specifications, including content balance. For TBTs they are then presented randomly to individual test takers. Then, accessing a test taker’s testing history, specific forms are not repeated until all the forms have been exhausted.

Prevent Item Scoring Key Disclosure. Do not provide the key to the answer to the questions. While this should be an obvious recommendation, many programs believe that a test should provide a learning opportunity for examinees, and provide specific feedback after each question. The feedback can be either simply whether the item was answered correctly or incorrectly, or it can also provide what the correct answer was. This feature in a high-stakes exam is inappropriate for many reasons. Not providing feedback keeps important information from test takers who may be inclined to share it with others.

Forward-Only Item Presentation. Green (1988: 79) suggests that “it would be better psychometrically to prohibit review” of test questions. To avoid the security problems associated with reviewing test questions, such as returning to a question to better memorize it, testing programs should establish a policy of forward-only item delivery. Once a question is answered and the next question is presented, a test taker would not be able to return to the earlier question(s). This policy has been implemented successfully for many large-scale programs, particularly those using CAT, including the ASVAB and the National Council Licensure Examination for Registered Nurses (NCLEX-RN).

Solutions for Inappropriate Retakes

The ability for a test taker to retake exams as desired is a security risk as several of the exam breaches in this chapter attest. Today, there is no reason why technology cannot be used to monitor the testing history of an individual and prevent unauthorized retakes. And it works this way for many programs. For example, it is a simple matter to check recent testing history and prevent the retaking of a test if the previous test had been passed (as in a certification or licensure exam). Even if the previous test were completed only minutes before, the system will be aware of the completion of that exam and can use that information in evaluating the retake situation. Many retake rules include the elapsed time since the previous exam. If a program does not allow a test to be retaken for five business days, the system can easily monitor and apply this rule.

Unproctored, unmonitored online exams without examinee authentication are those most at risk of suffering from fraud due to retake violations. When no authentication of a test taker’s identity exists, it is not possible to determine if a test
taker has taken a test earlier. The argument might be that the person has to provide his name and other demographic information, but an examinee can simply supply fake names or addresses, and retake the exam as many times as are necessary to see and learn what the test contains. When the examinee is sufficiently “prepared” with pre-knowledge of questions, he or she then signs up legitimately to retake the test.

**Prevention and Deterrence Solutions for Cheating**

Many of the solutions for theft described above will also work to prevent or deter cheating. For example, if test content is not stolen, it cannot be used to cheat. Those approaches will not be repeated here. Instead, this section will focus on additional strategies for preventing examinee cheating on technology-based tests.

**Reduce Testing Window Size**

The concept of a “testing window” refers to how long of a period exists where a test can be scheduled and taken. It is usually measured in days, weeks, or months. The longer the window the greater the opportunity to obtain test questions from earlier test takers or websites in order to cheat. There is no reported optimal window size, nor an industry-accepted rule for establishing it. Today, windows range from a few days to an indefinite length.

**Use Reliable Testing Center Networks**

Man of the security problems associated with “franchised” or “authorized” testing center networks can be avoided by administering a program’s tests through higher-quality networks. These proprietary networks are usually wholly owned by a test administration vendor, which has much more control over the quality of services offered. For example, proctors will be full-time, well-trained, and supervised. The operation of the center will be completely independent (i.e., no conflict of interest) from other purposes such as training or education. Equipment to support security will be enhanced, often including the most advanced biometric devices to authenticate the identity of examinees. Because of the greater professionalism, security is enhanced, and the service is suitable for any testing program. Of course, the service is a premium one and will be more expensive than lesser quality and less trustworthy networks.

**Delaying Score Reporting**

The US Medical Licensing Exam is administered under a policy that may delay the reporting of an individual’s test score (USMLE, 2008). The policy in part states:

> The performance of all examinees is monitored and may be analyzed statistically to detect aberrancies indicating that your scores may be indeterminate. In addition, evidence of irregular behavior may suggest that your scores do not represent a valid measure of your knowledge or competence as sampled by the examination. In these circumstances, your score report may be delayed, pending completion of further analysis and investigation …
Your scores may be classified as indeterminate if the scores are at or above the passing level and the USMLE program cannot certify that they represent a valid measure of your knowledge or competence as sampled by the examination.

This combination of analysis and decision-making is rare in high-stakes testing programs. Candidates, applicants, and other test takers have enjoyed the immediate reporting that is possible from technology-based tests. It is difficult to rescind a score once it has been reported. It is more difficult still to retract a certification, licensure, or admission decision, or a grade given after a course has been completed. These actions are made even more difficult if several weeks or months have passed since the exam was taken.

In addition to a policy similar to what the USMLE has created, testing programs are encouraged to adopt detection methods that can determine whether a particular score is aberrant or indicative of cheating in some other way (i.e., shows evidence of proxy testing or collusion). To retain the advantages of quick reporting provided by the use of technology-based tests, these detection systems must operate very quickly during the test or after the test has completed. If the detection system discovers and verifies a problem with a particular instance of test administration, it should relay that information immediately so that the reporting of the test score can be prevented and other actions taken. The ability to detect and report anomalies or verified instances of cheating is one of the capabilities and advantages of test administration using technology. It is surprising that few high-stakes testing programs have such a system in place.

Better Technology-Based Proctoring

Proctors and test administrators have enjoyed the benefits of increased technology used to enhance the efforts to monitor examinees during an exam. With closed circuit television technology, one or more cameras can be placed so that a proctor can view the entire testing center room without having to actually move around the room. He or she can monitor the examinees from an adjacent room.

More recently, Foster, Mattoon, & Shearer (2009) introduced a method of secure monitoring by requiring that each test taker have a webcam which allows a remote proctor the view of the examinee’s torso, arms/hands, and keyboard. The technology allows a proctor to be online as well, in the next room or hundreds or even thousands of miles away. In addition the new monitoring system allows the proctor to pause, suspend/restart, or stop a test in order to require the test taker to comply with testing rules. For example, if the test taker is using headphones, which are against the testing program’s rules, the proctor will pause the test and send a message that requires the test taker to remove and set aside the headphones. The message will also warn the examinee that the same infraction may result in suspending or cancelling the exam. According to Foster, Mattoon, & Shearer, immediate intervention by online proctors is very effective at stopping cheating or theft in progress, and at deterring those behaviors from the start. Since 2009, several organizations have introduced proctoring services that rely on remote proctoring capability. They differ in their use of different authentication hardware and methodology, and the degree of control proctors have over security incidents.
Proctoring methodologies such as those described by Foster et al. (2009) are assisted in the monitoring effort by a number of other Internet-based technologies. A few are listed here. Real-time data forensics analyses begin analyzing the examinee’s responses at the first question. Both specific responses made as well as latency of the response are included in the analysis. If enough forensics evidence of cheating is determined, an alert is issued and a proctor may monitor the particular examinee more closely to discover the reason. In some systems, remote proctors also have the ability to “rewind” the video of a testing session, to return to when the incident occurred as they continue to investigate. Online proctors can objectively apply rules set forth by the testing program to deal with different types and levels of cheating. For example, a different action may be applied to an examinee attempting to use Ctrl keys to copy text on the screen than to one using a textbook that was not an approved resource.

Results related to security issues from initial pilots of the new online proctoring methodology have been very positive. Foster (2009) reported the results for several distance education and certification testing pilots conducted in 2008 and 2009 in ten different countries. Almost all of the tests were administered in the examinee’s home. All exams were proctored from a single US location. Across such diverse tests for almost 6,000 students, the monitoring system and proctors detected security problems in less than 3 percent of the exams. All of the incidents were detected early in the test and were dealt with immediately by the online proctors. All tests with incidents were flagged and reported to the respective testing program for any further action.

Use Complex Item Types

The use of technology for testing allows the presentation of more complex items than the simple multiple-choice question. Some of the more complex assessments require the examinee to complete a task on the computer using a simulation of software or the actual software itself. Other item types provide lengthy case studies and the tools to search and review them. Still others ask the test taker to identify an area on a graphic by clicking on it using the mouse pointer or to drag objects from one place to another on the screen. Graphics, video, audio, and animations are used to supplement text almost routinely, whether the tests are Internet tests or computerized tests. Sireci & Zenisky (2006) provide a good review of new item types available for technology-based tests. Mulkey (2001) and Vaglio-Laurin (n.d.) have claimed that such performance-based items reduce security problems because they require the examinee to demonstrate the important skills. The latter states, “With (performance-based tests), even if information about an exam is shared, an individual still must be able to perform specific tasks under test conditions in order to pass.”

Certainly the use of such complex interactions during the exam makes it more difficult for many typical types of theft and cheating to occur. These new item types should be used for the security reasons mentioned, as well as to provide more accurate measurement of the knowledge and skills of interest. However, there remains a security risk: the use of such tasks and tools during the test reduces the number of skills that can be assessed during the time limits for the exam. In addition, the cost of developing such items results in further limiting the number of items created and used in the
Security Issues in Technology-Based Testing

exams. The reduced amount of content may under-sample the domain of interest and allow cheaters to specifically learn of and practice only those skills measured on the exam. This was recognized by researchers looking into performance-based testing issues for the health professions. Swanson, Norman, & Linn (1995) reviewed the various performance testing models in the health fields and reported several important “lessons learned.” One of these, Lesson 6, states, “Because performance-based assessment methods are often complex to administer, multiple test forms and test administrations are required to test large number of examinees. Because these tests typically consist of a relatively small number of independent tasks, this poses formidable equating and security problems” (pp. 9–10).

Stronger Authentication and Reauthentication

Technology-based tests match up well with new hardware-based and software-based biometrics. The use of fingerprint and palm readers has helped in the process of authentication, supporting or replacing the use of forms of government-issued identification documents. Because these documents can be easily faked and have other security issues, new biometrics are necessary. In a move more tuned to Internet-based tests, the field of keystroke analytics has provided technology that can establish a person’s typing patterns as a biometric, and can use that information to authenticate him or her before and during exams. During an initial enrollment phase, typically occurring at the point of registration, examinees may be asked to type a phrase 10 to 15 times. Information relating to how long a person dwells on keys and the amount of time to move from key to key is combined into a single quantitative value. This calculated value is unique for each examinee, even if each of them is typing the same phrase. When a test is to be launched, the typing of the phrase for the particular examinee must match the typing during the enrollment phase in order to access the test. It is very difficult, if not impossible, for a proxy test taker to operate, even in collusion with testing center administrators, because the biometric system will not allow the test to launch unless the phrase is typed properly.

In real time, another digital technology can be used to determine whether a test should be launched or not. A photograph taken and stored at some early enrollment stage is compared to a digital photograph taken just prior to the scheduled launch of an exam. Online proctors with no motivation to collude with the test taker can view both photographs and decide if they match. If they are judged as matching, the proctor can authorize the launch. The photographs are stored for later review if necessary. Successful use of this photo-based biometric technology, combined with keystroke dynamics, has been demonstrated and reported by Foster (2009).

Remove Conflicts of Interest

Technology-based testing center networks were set up in the early 1990s by contracting with training organizations to administer tests to the same students they were training. It was expected – and that expectation was reinforced by contract language – that those centers would keep the testing effort completely separate from the training effort. However, often that is not the case. Today, tests are given in places and proctored by individuals who have a vested interest in how the examinees
perform. Although most proctors are undoubtedly honest, the industry has been hurt enough by having stakeholders double as proctors. Programs have the ability and responsibility to prohibit testing at locations for which the test administrators have a conflict of interest.

Detect Proxy Test Taking

As discussed in a previous section, organized proxy test taking is one of the most serious security threats facing the testing industry. Impara, Kingsbury, Maynes, & Fitzgerald (2005) reported on a case study of proxy test taking, for purposes of better understanding how it can be identified and/or prevented. One pattern they observed is that proxy test takers will often take the same exam for different students in immediate succession (i.e., one right after the other). Many of the peripheral statistics on these exams are very similar, such as the number correct score, the length of time to complete the exam, and the pattern of which items are answered correctly or incorrectly.

Fortunately, this constellation of data provides clues as to how one might prevent organized proxy testing from occurring. The best place to prevent such testing is at the testing center. A trusted and competent administrator or proctor would not allow a person to take the same test many times using different names. Proper identification procedures would have quickly revealed the attempt at fraud. However, administrators and proctors have been known to occasionally collaborate with proxy testing organizations to facilitate their getting access to tests. One safeguard against this is to use biometrics (e.g., fingerprinting, palm vein reading, retinal scans, etc.) to help authenticate the test taker. Automated biometrics, especially if combined with better human vigilance, and stronger testing program policies, would have indicated that the actual candidate or student was not present at the scheduled time and would not have allowed the test to launch. Having such a practice in place would make it virtually impossible for the proxy test taker to fulfill his or her purposes.

Use of Communication Devices During Exam

Cell phones, pagers, and any other communication devices present a very serious threat during exams, because they allow secure test information to be captured and transmitted to off-site locations across the world, and because they enable test takers to discuss test content with and receive help during the exam from others, either within or outside the testing room. Because of the dangers associated with allowing candidates to use communication devices during exams, it is important to take steps towards preventing their use. Preventing use of communication devices in any situation can happen by disabling the wireless network, confiscating the device prior to the exam and returning it at the end of the exam, or by using technology to jam the communication signal at the testing location. None of these is without problems. For the first, other public or personal networks may be available. For the second, the test taker might have a second device, may not disclose having a phone, and may resist a search. In addition most testing staff may not assume the responsibility of keeping the devices. For the third, jamming communication signals is illegal in some countries and represents a risk if an emergency message needs to be sent or
received. Examinees should be informed that these devices are not allowed and that they should either not bring such devices to the testing location or should place them in storage lockers, if available at the site.

**General Deterrence Solutions**

In general, deterrence will be effective if the test taker has the perception that he or she will be caught and punished severely for test fraud. And when security problems are detected, the program must make every effort to address the problem or breach, however minor. But general approaches to deterrence are equally important to the need to inform the examinee population and the public of the efforts to protect the exams. This disclosure, which should be prudent to the point of not disclosing proprietary or secret methods, will discourage or inhibit many from engaging in attempts at test fraud.

**Announce Prevention and Deterrence Activities**

All testing programs, whether technology-based or paper and pencil, should communicate in general terms the consequences for cheating and sharing test information, and the strategies that will be used to detect such efforts.

It would be effective to announce procedures that have been instituted to stop a particular activity. For example, new biometric tools, such as palm pattern readers, are now being used to detect proxy test takers.

If an internet-based testing program has a lockdown program that detects if a person attempts to access a computer resource, such as the Internet, it should make that fact known to the testing population. If there are sanctions applied, those should be clearly stated as well. For example, it should be disclosed that the program can detect if a person attempts to use the keystroke combination Alt-Tab, used to switch from one software application to another in the Windows operating system. Furthermore, the disclosure should state the consequences if a person attempts to use Alt-Tab, which might include suspending the test immediately until a security review committee can evaluate the incident. The disclosure of the ability to detect Alt-Tab and the consequence may be sufficient to significantly reduce attempts to use Alt-Tab during the test.

**Create Strong Security Policies**

Examinees should know the rules regarding theft and sharing of content. This should be part of a clearly written policy agreement which examinees are expected to accept prior to taking an exam. Technology-based exams make this process easier by providing it in electronic form at some point before the exam. If the agreement is not accepted the test would simply not be available to the test taker. Examinees should also agree by contract not to cheat during the exams. It should be clear to them that this means the use of pre-knowledge, help by instructors/teachers, use of inappropriate aids, or the use of proxy testing services. It should be clear what the penalties would be for cheating or theft of questions. The completed agreement will be important when taking action against violators.
Follow Through Completely Based on Policies

While not unique to TBT, this recommendation is sufficiently important to mention in multiple chapters in this Handbook. Once a breach is discovered and the subsequent investigation has uncovered individuals or organizations responsible, the testing program should follow through completely with legal action, or business sanctions, or other consequences. These all depend on the security policies established, which will be agreed to by partners, vendors, test takers, etc. A complete set of potential actions based on the security breach action plan will be very effective in deterring test takers and others from cheating on or stealing exams. What has been agreed to and the possible actions should be well publicized.

Controlling Item Exposure and Over-Exposure

Exposure of test questions is a good thing. It occurs naturally as tests are administered. Displaying the contents of a question to a level sufficient for the test taker to read, understand, and complete those actions required is a necessary feature of all exams. If individuals never tried to memorize them and share with others, or tried to capture them digitally with a camera or cell phone, or if questions were not displayed again when a test is repeated to the same examinee, the exposure of questions would never be considered a security issue. Theoretically, questions can be displayed thousands, hundreds of thousands of times, or even millions of times, as long as they are displayed only once to each examinee, and that examinee does not have pre-knowledge of the question. However, as has been discussed, item pre-knowledge is a serious concern. These concerns are perhaps most serious for CAT administrations, because these tests tend to be shorter and some questions are more heavily exposed. Consequently, much CAT research has focused on reducing the likelihood of examinees having prior exposure to the questions by controlling item exposure levels through the use of special algorithms for item selection (see McBride & Martin, 1983; Chang & Ying, 1999; Chen & Lei, 2005; Chen, Lei, & Liao, 2008; Davey & Parshall, 1995; Hetter & Sympson, 1997; Parshall, Davey, & Nering, 1998; Stocking & Lewis, 1998; Yi & Chang, 2003). Although controlling item over-exposure is an important security consideration, the challenge with imposing restrictions on item exposure is that, unless the best, most informative items are administered to candidates, the efficiency of the CAT algorithm suffers (i.e., more items must be administered to achieve a score with the same level of precision). The various item exposure control algorithms allow test publishers to place limits on the frequency with which individual items are to be selected for administration, while simultaneously trying to minimize the loss of efficiency and maintaining the appropriate content balance. Discussing the relative merits of the different item selection methods within CAT is beyond the scope of this chapter; for a recent and comprehensive review of these methods see Cohen & Wollack (2006).

Way et al. (2002) implement a somewhat different strategy to control item over-exposure and focus more on maintaining and replenishing a pool of items. They propose a set of rules for item usage, including one that states, “Any item exposed to more than 15,000 examinees must be retired” (p. 155). They admit that there is no empirical basis for the rule, but feel that it is “conservative.” An assumption underlying this rule and similar ones proposed is that exposure occurs gradually over time, as the same items are presented to more and more examinees. At some point the
item characteristics will begin to change. The goal is to implement a strategy that calls for items to be retired and replaced before those characteristics change substantially (see Han & Hambleton, 2008).

Exposure control is both a science and an art. An item exposure rate of one is too many if that one person captured the question with a digital camera and put it on the Internet. In fact, a single exposure is not even necessary if the file containing the questions was captured before the first test was taken. If the item was shared to all on the Internet via a braindump site then it is compromised and it is doubtful that the item will function as intended. The concept of “item shelf-life” is not defined by number of naturally occurring presentations of the questions, but by how quickly the question can be stolen and shared, and how extensive that sharing is. It is more affected by security measures in place to protect it than the number of presentations to examinees, so can vary widely program to program. A more useful definition of item exposure might be the number or percent of people with pre-knowledge of the item before taking the test.

One way to measure pre-knowledge would be to use the Trojan Horse methodology (Maynes, 2009) discussed earlier, which indicates at the test level how many people have pre-knowledge of the items. Han & Hambleton (2008) describe several item compromise detection algorithms as they analyze continuous item p-values to determine if the item is performing at the same level of difficulty as it did when it was first presented. Item p-values are continuously calculated as the item is presented in more recent exams. The recalculated p-values are plotted on a chart with upper and lower limits. When the p-value increases (item gets easier) due to exposure and crosses the upper limit, it is time to replace the item with a non-compromised one.

One must remember that the exposure of questions is natural and normal. Efforts should be made to (a) protect the questions from theft by stronger security efforts, including imposing exposure control algorithms, and (b) run continuous forensics efforts to evaluate the degree of compromise. Once compromise is detected, the item is over-exposed and should be replaced. Technology-based tests, at least compared to paper-and-pencil tests, allow for replacing items and republishing new tests in a relatively straightforward and quick manner. Written policies and examinee agreements should reinforce the testing program’s right to make such changes without notice.

Conclusion

We need to remember that the cheating/theft war is one that will never be over. We cannot win in the absolute sense that cheating and test theft will be completely eliminated. When we win a battle, the cheaters and thieves will create new threats and try again, usually successfully until the testing industry can, in turn, develop prevention, detection, and deterrence strategies. What we can do is use practical security approaches and the technology tools that are available with technology-based tests to stay ahead of the game, perhaps even being able to anticipate what the next threats will be. If we cannot prevent or deter those threats, we can effectively detect when they occur and respond quickly to reduce or eliminate their effects. There is ample evidence included in this chapter and in the records of professional and determined high-stakes testing programs to demonstrate that security success is attainable.
Note
1. A main reason why some Internet testing services continue to use the download model is mainly to continue to administer test questions if the Internet connection is disrupted.

References
Security Issues in Technology-Based Testing


