INTERNATIONAL TEST COMMISSION

The ITC Guidelines on the Security of Tests, Examinations, and Other Assessments

6 July, 2014, Version 1.0

Final Version

Document reference: ITC-G-TS-20140706

Formally adopted

The Council of the International Test Commission formally adopted the guidelines at its July 2014 meeting in San Sebastian, Spain.

Published online

This document was officially published online after the General Meeting of the ITC in July 2014 in San Sebastian, Spain, and can since be found online on the ITC website at http://www.intestcom.org.

Published in print

This document was not yet published in print.

Please reference this document as:

International Test Commission (2014). International Guidelines on the Security of Tests, Examinations, and Other Assessments. [www.intestcom.org]

ACKNOWLEDGEMENTS

These guidelines were prepared under the leadership of Dr. David Foster, Kryterion, Inc. and Caveon Test Security (USA), with support from Eugene Burke, SHL (UK), and Casey Marks, Cambridge Assessments (USA). The standards have drawn upon a variety of papers and publications in the field of test security, and we would therefore like to thank the various contributors in this field as well as those who have personally assisted in developing the standards contained in this document. Specifically, we would like to thank the contributions of the following:

David Bartram (United Kingdom)
Ian Coyne (United Kingdom)
Dragos Iliescu (Romania)
Tom Oakland (United States of America)

The author also acknowledges the effort and valuable comments and suggestions of a number of members of the community who have contributed in the review phase of the document: Sara Gutierrez (on behalf of CEB SHL Talent Measurement), William G. Harris (on behalf of the Association of Test Publishers), John Hattie, John Kleeman (on behalf of Questionmark), Fredi Lang (on behalf of the Diagnostics and Testing Advisory Board of German Psychologists' Association), Peter Macqueen (on behalf of the Tests and Testing Reference Group of the Australian Psychological Society), Marcus Scott (on behalf of Caveon Test Security), Richard Smith (on behalf of the British Psychological Society).

We would also like to acknowledge a number of key standards and guidelines that have assisted us in developing the materials you will find in this document. These include:

- American National Standards Institute (ANSI) (2006). *Guidance for Conformity to ANSI/ISO/IEC 17024: Requirement for Certification Program Security*.
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (1999). *Standards for Educational and Psychological Testing*.
- Caveon Test Security (2009). *Test Security Standards*.
- Association of Test Publishers (ATP) (2002). *Guidelines for Computer-Based Testing*.
- International Test Commission (ITC) (2005). *International Guidelines on Computer-Based and Internet Delivered Testing*.
- National Council on Measurement in Education (NCME) (2012). *Testing and Data Integrity in the Administration of Statewide Student Assessment Programs*.
- National Organization for Competency Assurance (NOCA) (2001). Certification Testing on the Internet.

In addition, we would like to recognize the contribution of the following reference text devoted entirely to protecting tests and assessments:

- Wollack, J. A. & Fremer, J. J. (2013). *Handbook of Test Security*. New York: Routledge.

SUMMARY

The amount and severity of security threats have increased considerably over the past two decades, calling into question the validity of assessments administered around the world. These threats have increased for a number of reasons, including the popular use of computerized and online technologies for test administration and the use of almost undetectable technologies for capturing test content and illegally sharing it instantly across borders and cultures. No assessment program, large or small, is immune to this potential damage.

The International Test Commission has recognized the critical need for every organization with an important assessment program to be aware of these and prepared to counter them. It was for this purpose that these guidelines were developed. Knowing the threats and the guidelines will lead to effective measures to protect the program and its assets, maintaining the value of the tests and assessments to the international community.

The guidelines listed in this document provide recommendations on planning for better security, maintaining security during the development of tests and while they are administered, and responding well when a security breach occurs. Following these guidelines will create a significant protective barrier between those who willingly commit test fraud and the valuable assets a program has spent time and money to build.

CONTENTS

INTRODUCTION

The purpose of the International Test Commission (ITC) Guidelines for Security of Tests and Other Assessments

The need to secure tests, exams, and other forms of assessment has increased in importance along with the growth of testing as well as the increasing role of technology in deploying, administering, and scoring tests, notably via the Internet.

All stakeholders in the development and use of tests would agree that the value of a score from a test or other structured assessment is diminished when it is subject to some form of cheating or test theft. Cheating is defined as any attempt to improve the score on a test, examination, or assessment by fraudulent means. Test theft is defined as an attempt to steal test content before, during or after its intended use.

The principal purpose of these guidelines is to share key elements of best practices through which test developers, test sponsors, testing service providers and test users can promote the security of their testing and assessment programs and defend the value of the information provided by the scores obtained from these programs.

Cheating, test theft, and other breaches may happen to even the most conscientious of programs. However, an active security management program will help ensure breaches are fewer and their damages are limited.

Audience for the Guidelines

There are many stakeholders engaged in the testing and assessment process. Each may be affected by security breaches and may benefit from knowledge and application of these guidelines. Seven stakeholder groups are described below.

- **Test Takers**. These individuals are personally completing a test or are being evaluated in other ways. These persons also may register, pay for, and schedule the test.
- **Test Developers**. These are individuals or organizations that are responsible for the design and creation of the test or assessment. These may be part of a service provided by others.
- **Test Administration Service Providers**. These organizations have technology and distribution channels (e.g., testing centers) to make sure that a published test is available at times and locations convenient for the Test Takers.
- **Test Security Service Providers**. These providers offer specific security services (e.g., forensic analysis) to enhance security efforts. Test security service providers may be part of or separate from a larger service organization.

- **Technology Service Providers**. These organizations provide various services to other stakeholders, including but not limited to database services, item banking technology, communication services, adaptation services, and storage support.

- **Test Publishers or Owners**. These organizations or persons own the content of a test and authorize its use for specific purposes. They also may contract with service providers as needed.

- **Test Users**. The Test Users are stakeholders who make use of test information, including scores, for individual or group decisions or for policy-making.

How the Security Guidelines are structured

These guidelines are structured around key actions that underpin effective test and assessment security. These actions have been classified as follows:

- **Developing and implementing a security plan** that outlines necessary preparation, including the creation of a security incident response plan, and sets out policies and procedures for actively managing security;

- **Implementing security for the testing and assessment process** that covers both test/assessment design and development as well as the administrative procedures for deployment of tests and assessments; and

- **Responding to security breaches** when cheating or test theft has been discovered.

How to put these Guidelines into practice

These guidelines are intended to be applicable internationally. Many contextual conditions may affect how guidelines are managed and realized in practice. These contextual conditions must be considered at the local level when implementing these guidelines in any location. These conditions include:

- Social, political, institutional, linguistic, and cultural differences between assessment settings;

- Laws, statutes, policies, international standards, and other legal documentation that address testing issues;

- Laws applying to the various countries through which test data may pass, be stored, or be used; and

- Existing national guidelines and performance standards set by professional societies and associations.

THE GUIDELINES

Scope of the Guidelines

The incidence of test fraud is greater in high-stakes scenarios where a test, exam, or structured assessment produces a score that has significant consequences for the test taker[1] and/or other stakeholders. Such scenarios include educational tests taken to obtain admission to an educational program or taken during and at the conclusion of a program to obtain a qualification. In clinical settings, such scenarios may include decisions regarding clinical treatment or legal procedures that result from a diagnosis of a test taker. In employment settings, such scenarios include obtaining employment and promotion within an organization. Allied to employment settings are those related to skills assessments through which test takers are awarded professional qualifications such as certifications or licenses. In forensic settings, such scenarios may include determining the ability to be tried for a crime and, if convicted, the severity of the sentence.

While these guidelines focus on test use, given the maturity of the research and practice in test security, the examples provided in the previous paragraph show that security is an issue that may apply to any structured assessment used to evaluate the knowledge, skills, abilities and psychological attributes of individuals. For example, in an employment setting, a test taker may be assessed through an interview for which test takers could be prepared through coaching and have access to typical interview questions. Behavioral observations in the workplace or classroom constitute another form of assessment that can be affected by security threats, particularly if the observer has an interest in the outcome of the observation. Although the terms *test* and *exam* are used more frequently, the reader will find in these guidelines information that may help improve security for all types of assessments.

Assessment intended to be lower stakes (e.g. a 360 appraisal used to identify training and development needs of employees) may rise to a higher level in the eyes of test takers when they recognize the consequences (e.g. access to training and development programs and subsequent eligibility for rewards, including salary increases and/or promotions) that are dependent on such assessments. Stakeholders are likely to find the principles presented in these guidelines to be of value to them irrespective of the intended importance of the assessment. Recognizing this general value, these guidelines do not apply to contexts that do not require test security, such as self-assessments and practice tests.

These guidelines frequently reference technology to help prevent or detect test fraud. While test administration increasingly has moved to the use of computers and the Internet in many settings, test security is an issue that applies to any form of test administration or assessment. Therefore, the principles described in these guidelines apply equally to paper-and-pencil or manual administrations, to technology-based modes of testing and assessment, and to hybrid models where more than one administration mode is used.

[1] Test taker and examinee are terms that are often used interchangeably in reference to persons who sit for a test, examination or assessment, whether the stakes are high or low. In these guidelines, we use test taker and examinee to denote persons who sit for any type of assessment irrespective of the purpose or stakes involved.

In short, the scope of these guidelines is to promote security for all tests and assessments irrespective of whether they are deployed in high- or low-stakes scenarios, and to promote best practices, whether tests or assessments are delivered manually or electronically, recognizing that differences in the methods and levels of security imposed may vary depending on the type and/or setting of the assessment.

Security is not all-or-nothing. What is generally the case is that a balance needs to be drawn between the risks of cheating and theft on the one hand and the costs of preventing it on the other. This balance depends on the stakes involved. These guidelines are intended to cover what can be done to maximize security but recognize that not all these guidelines are necessary to implement in all cases. This emphasizes the need for a risk analysis to be carried out for each new scenario and security measures put in place that address and mitigate those risks. It also emphasizes the need to look at the security plan for the whole assessment process.

The Guidelines are divided into three parts: (1) Developing and Implementing a Security Plan, (2) Implementing Security for the Testing and Assessment Process, and (3) Responding to a Security Breach. Each of these parts is presented in that order below.

Part 1: Developing and Implementing a Security Plan

Basic terminology for a successful security effort includes the concepts of threats, risks, vulnerabilities, and breaches. The preparation of a successful program requires knowledge that specific security threats exist and that they can be associated with an estimated amount of risk. As an example, vulnerabilities or weaknesses in the program security as well as inadequate personnel training increase the level of risk. Risk can be calculated informally, given circumstances at a point in time, such as:

- the likelihood a threat will be successful,
- the ease with which program vulnerabilities can be exploited,
- the amount of damage a threat may cause if it becomes a successful breach, and
- how prepared the program is to detect/stop a breach and repair the damage.

An actual example may be helpful to illustrate these concepts. Given the high-stakes nature of statewide testing in the US, there have been a large number of investigated security incidents involving some school administrators and teachers alleged to be manipulating test scores (the threat) by changing answers sheets, coaching students, providing access to students before the test, and in other ways (e.g., cheating methods, see below). The likelihood that this cheating will actually happen and cause the expected damage (the risk) can be analyzed in advance by looking at the prevalence of breaches in other states, the damage those breaches have caused, and understanding that teachers and administrators are the individuals actually responsible for test administration (a vulnerability). A breach occurs when cheating actually occurs and is detected.

Using a risk analysis process, and considering its organizational goals, a program can prioritize how to use its limited resources to remove or reduce threats, strengthen the vulnerabilities,

implement detection mechanisms to quickly discover attempted or successful breaches, and prepare to minimize and remediate the effects of a breach.

The establishment of an effective security plan requires one to understand the nature of the program's current security threats and the risks associated with them. A security threat is a source of either potential cheating or test theft. For example, a cheating threat exists when a mobile phone is available to be used to receive text messages during a test. A test theft threat exists when someone is able to access a storage device or location and capture some or all of the content of a test. The need to establish and revise an effective security plan increases as the particular threats and risks for a testing program are better understood. A properly developed and managed security plan will reduce the threats and mitigate damage from breaches.

Table 1 and **Table 2** list cheating and test theft threats, respectively (Foster & Miller, 2012).

Table 1. Categories of Cheating Threats

Cheating Threat	Description
Using test content pre-knowledge	Test taker obtains actual test questions from a reliable source prior to sitting for the exam.
Receiving expert help while taking the test	Test taker receives help from a teacher or other confederate during the test.
Using unauthorized test aids	Test taker uses unauthorized aids during a test, such as cheat-sheets, cell/mobile phones, headphones, programmable calculators, etc.
Using a proxy test taker	Test taker uses a professional proxy testing service or simply has a friend or colleague take the test.
Tampering with answer sheets or stored test results	Following the completion of a test, a person (e.g. a teacher) may tamper with answer sheets, changing wrong answers to correct ones. Alternatively a test scoring database can be entered in order to raise test scores.
Copying answers from another test taker	Test taker copies answers provided by another test taker during a test.

Table 2. Categories of Test Theft Threats

Test Theft Threat	Description
Stealing actual test files or booklets	Test content is most vulnerable to theft at particular stages of test distribution (e.g. when files are stored on a server or test booklets are kept in a storage room). Inadequate access controls allow thieves to capture entire test content along

	with answers.
Stealing questions in test through digital photography or copying devices	Test questions can be captured as they are displayed during a test. A thief may use a hidden or otherwise undetectable high-resolution digital camera or other copying devices (e.g., pens that scan).
Stealing questions by recording test content electronically	For technology-based tests only, an entire test session, including all test questions, can be captured with an automated procedure by using a digital recording system connected to one of a computer's output ports.
Memorizing test content	Test taker memorizes questions to be recalled and recorded at a later time. As part of an organized effort, this kind of theft is termed "harvesting."
Transcribing questions verbally	Oral or written content may be captured during a test. This may involve the use of audio and text recording devices, such as cell/mobile phones, two-way radios, or notepads/scratch paper.
Obtaining test material from program insider	An employee or contractor of a testing program may steal test content during the course of test development, publication, or distribution.

A risk analysis evaluates the likelihood of the threats listed in Tables 1 and 2 being successful together with the amount of damage they could cause if the breach were successful. Two examples follow.

A single person cheating by his or her own effort is likely, even common, for any testing program. Its damage generally is limited to a single inaccurate decision made because of a single inaccurate test score. On the other hand, a stolen and online-distributed test booklet may inappropriately increase thousands or tens of thousands of test scores. While this event is less likely, it results in much greater damage. An organization must decide the degree its limited resources should be applied to detect, deter, and deal with individual cheaters or to establish procedures that make stealing and distributing test booklets more difficult.

Tables 1 and 2 present a taxonomy of the few threat categories that are known today. However, for each category, the number of actual *methods* that people use to cheat or steal may number in the hundreds. Following the lead of the banking industry, a comprehensive security effort should use multiple layers of security procedures, given the well-established assumption that several methods, working in concert, will be more successful than a single method. These guidelines are intended to be used in concert with each other and to provide specific guidance as a program prepares to address its security risks effectively.

Specific Guidelines on Developing and Implementing a Security Plan

A comprehensive document outlining the security plan is necessary to manage the integrity of all test and assessment materials, as well as test scores and the decisions based on them.

1. This document should identify the security roles and responsibilities at all key stages of the process from design and development through deployment, results collection/storage/analysis, and distribution/administration. It may cover some or all of the following roles:

    a. Security Director. When possible, a program should appoint a security director who is responsible for all aspects of the program's security.

    b. Security Committee. A program should form a security committee, chaired by the Security Director, of individuals responsible for the creation/maintenance of the security plan, evaluating severity of and applying policies regarding security incidents, overseeing responses to security breaches, and working in other ways to create and maintain a viable security plan.

    c. Managers. Individuals with responsibility for test development, test administration, and test results collection and storage should be trained regarding and adhere to the security policies and procedures set forth in the security plan.

    d. Proctor, Invigilator, or Test Administrator. When used as part of the security process, these individuals are responsible for the secure administration of the test, including the authentication and vigilant monitoring of the test taker throughout the testing session. Proctors and/or test administrators should not also serve as instructors, subject matter experts, trainers, or in other roles that provide access to content assessed by the test or in other ways present possible conflicts of interest that may impact a test taker's performance.

    e. Test Security Services Providers. These individuals assist the program in identifying vulnerabilities, helping to prevent security problems, detecting breaches when they occur, determining the extent of damage, recommending courses of action, and perhaps carrying out that action. These security professionals include consultants, investigators, data forensics analysts, web monitoring specialists, legal experts, and others.

2. The security plan document should specify the rights and responsibilities of test takers in taking the test or assessment and how the test taker's acknowledgement of those rights and responsibilities will be recorded.

    a. Test takers have the right to take high-stakes assessments that are secure, so that no other test taker gets an unfair advantage due to cheating or other forms of test fraud.

    b. Test takers suspected or accused of test fraud have the right to due process.

    c. Test takers have the responsibility not to disclose test content to others and to report such activity when discovered.

3. The security plan document should be available to stakeholders upon request.

4. The security plan should include a breach action plan that outlines what to do in the event a security breach occurs. The action plan should include goals, timelines, key personnel, reporting systems, an escalation path, public disclosure rules, media relations, and specific corrective actions to be taken depending on the nature of the incident or breach. Among others, corrective actions may include sanctions for offenders, cancelled or invalidated scores, re-testing procedures, replacement of item pools or test forms, and legal action.

5. Security rules should be indicated clearly in the security plan and communicated to all interested parties. Consequences for violations of those rules should be clear.

6. The security plan should be approved by the appropriate set of stakeholders and reviewed at least annually.

7. The security plan should document the security requirements for the Information and Communication Technologies (ICT) policies and procedures for employees, contractors, and all service providers. These requirements will discuss the secure storage of and access to test content, test results, other testing information, test taker information, and the protection of that information during communication and data transfer processes.

8. The security plan should include references to privacy laws for different countries and regions where testing occurs. The plan should indicate how policies and procedures are modified to accommodate these differences. Efforts to protect the data of individuals and organizations must be consistent with applicable laws and policies.

9. Sufficient funds should be available for implementing security prevention and monitoring activities outlined in the security plan document. In addition, a reserve fund should be established sufficient to respond to the most serious of potential security breaches. The security budget should be reviewed regularly, adjusted as needed, and be commensurate with the identification of new threats.

10. Security training materials related to the roles and responsibilities outlined in the security plan document should be created and provided to all individuals involved in the testing enterprise.

11. Non-disclosure and other agreements should be executed routinely for all parties, including test takers, service providers, and program employees. These agreements will require acknowledgment of the copyright and ownership of test and assessment content, acknowledgment of acts considered to be fraudulent, and the potential consequences of such acts. The agreements will require individuals to acknowledge they will not disclose the specified proprietary information.

12. The test owner should copyright or otherwise legally establish ownership to protect its test content in countries where the tests will be administered.

13. The security procedures of all service providers should be monitored and audited periodically to evaluate the effectiveness of policies and procedures. Internal or external security experts can perform this service.

Part 2: Implementing Security for the Testing and Assessment Process

After developing an approved security plan, security for processes that occur before, during, or after a test or assessment can be designed, created, implemented, and managed. Important steps in the processes that have security implications include:

- Test taker registration
- Authentication or identification of test taker
- Test and item design
- Test development
- Test publication and distribution
- Test administration
- Test scoring
- Test results and candidate information collection and long-term storage

Many of these include processes that require the management and distribution of sensitive materials between stakeholders.

Specific Guidelines on Implementing Security for the Testing and Assessment Processes

1. Test takers should be required to register formally for an assessment. Registration for the administration and scheduling of a test or assessment should include, at minimum, the issuance of a specific and unique username and login or password for each test taker.

2. At the time of registration or scheduling, test takers should be informed that they will need to follow proper authentication procedures. The test takers may be known to the testing organization or have organizational IDs. Alternatively, the test takers should be asked to provide verifiable information, such as legally-recognized, government-issued documents with photographs, or be required to participate in biometric authentication procedures. In some testing procedures, it will be necessary to inform test takers of the need to provide authentication at some point following the test session (as in the use of screening tests in pre-employment assessment)

3. Registration procedures help ensure only qualified persons register for and schedule a test or assessment. Requirements for qualification may include the completion of a course of study, taking and passing a pre-requisite test, or paying a fee. If applicable, the requirement may also include a length of time that must elapse before a test can be re-taken.

    a. If permitted and compatible with privacy laws, a list of "restricted" high-risk test takers may be created and maintained by the testing program that informs both offline and online

registration/scheduling systems to prohibit or limit testing of those individuals according to established program rules.

4. Re-testing policies should be developed to reduce the opportunities for item harvesting and other forms of test fraud. For example, a test taker should not be allowed to retake a test that he or she "passed" or retake a test until a set amount of time has passed.

5. Registrations for a test or assessment by test takers should be closely monitored to prevent them from taking the test more often than allowed in order to minimize opportunities to harvest items.

6. Test designs that limit item exposure or change the order of items, while maintaining psychometric quality, should be considered. These designs include how items are selected and presented (e.g., computerized adaptive tests, linear-on-the-fly tests, multi-stage tests, multiple equivalent forms), whether items can be marked for later review during a test, and early stopping rules.

    a. For some types of tests, when enough questions have been presented and answered to produce a score with an acceptable level of reliability and to provide evidence of validity (e.g., content coverage) for decision making, the test design may terminate item presentation, thus preventing the unnecessary exposure of additional questions.
    b. The presentation of questions could also be designed to end or be modified in some way if there is evidence that a test taker is unmotivated, cheating, stealing questions, ill, fatigued, or for some other reason is not able or willing to provide an accurate estimate of the attribute being assessed.

7. Programs may consider a forward-only item presentation design that does not allow test takers to gather or amass items for possible capture (e.g., through memorization or digital capture). Some test and item designs are more secure when they restrict or reduce the ability to mark items and later review them (e.g., computerized adaptive tests).

8. The exposure of test or assessment content should be actively monitored and controlled. For example, developers should create assessments where selecting items from an item pool does not result in an unplanned and unmonitored over-exposure of items.

9. Larger item pools may support test administration procedures that improve control and management of item use and exposure.

10. Items should be designed to manage and perhaps limit the exposure of content. There are many ways to do this. Here are a few examples. Please note that changes to existing test development systems, delivery platforms, and database storage systems may be needed to accommodate new item types and the use of alternative formatted items (e.g. forced choice formats for self-report items, simulations, uses of multimedia, drag-and-drop interactions).

    a. When using multiple choice formats, consider not showing all options (for example, one variety presents options one at a time until the question is answered correctly or

incorrectly, or another variation would be to pull options from a larger pool, both of which only present a subset of available multiple choice options).

b. When using multiple choice formats, options can be presented randomly to confuse test takers that may have prior knowledge of a test.

c. The use of video, audio, simulations, and other forms of media may make it more difficult to capture test content and may prevent some cheating methods.

11. The use of a subsequent follow-up or verification form of the test or assessment may be required to confirm the score results of a prior test administered under less secure conditions. This verification process should be undertaken with a test taker's knowledge and consent.

12. Stringent statistically-based time limits should be established to provide adequate time to complete a test while reducing opportunities for the use of cheating aids and the theft of test content.

13. Test content must be carefully protected during the development stage as items and tests often are sequenced through a number of steps where psychometricians, editors, subject matter experts, and others have required access to them.

    a. The items and tests must be protected by limiting exposure to only those individuals required to author or review them, and then for only a limited time.

    b. Strong access procedures should be employed (e.g., usernames and passwords, or biometrics).

    c. Those who have access to test content and tests should be subject to background checks and strong non-disclosure agreements.

    d. Items and tests forwarded for review to other servers and temporarily out of immediate control should be removed or destroyed immediately at the end of the review and after the changes have been collected. That removal or destruction should be verified.

    e. The ownership of items must be established (e.g., copyrighting) according to country specific regulations and policies.

    f. Individuals involved in the development process should be trained in how to recognize and report security breaches.

14. Tests should be protected during production, publication and distribution.

    a. Servers used to hold testing content should be housed in a professional data center certified to international standards (e.g., ISO 27001 or SSAE 16) and using ICT security measures (e.g., firewalls and intrusion detection).

b. Individuals who prepare these tests should be trustworthy and be required to complete non-disclosure agreements.

c. When test content is distributed, whether in booklets or digital file form, it should be protected at every step of the distribution process and stored securely at testing locations. Technology-based test administration service providers should track releases of and apply security patches promptly to authorized operating system and application software.

d. Digital content should be protected by strong encryption schemes, whether that content is sent in its entirety to be downloaded to a remote server or sent item-by-item in real-time during an Internet test.

e. Test content that resides for any length of time on a server at a testing center must be protected at all times by strong user access controls (e.g. usernames and passwords) and strong encryption schemes.

f. Tests should remain at testing locations for the minimum amount of time, governed by program and test administration policies.

g. When a test no longer is needed at a testing location, the content should be removed and/or destroyed. Its removal or destruction should be verified, and the content should not be recoverable.

h. Developers should ensure that the distribution of all sensitive materials is documented clearly and can be traced, including the return of and/or destruction and disposal of materials, if appropriate, after use. Their return or destruction should be verified.

15. Tracking methods (e.g. paper or digital log sheets) should be used to record periods of control, access and changes to files.

16. Test takers should understand the security rules and consequences for their violation before registering and scheduling tests.

    a. Test takers should be made aware, well in advance of a test event (e.g., perhaps as part of an honor code or ethics agreement) that they will be required to read, acknowledge, and agree to abide by security rules prior to the start of a test.

    b. The consequences of breaking security rules should be clear.

    c. Test takers should have the opportunity to either agree or not agree to those rules before a test launches. Test takers who do not agree to the rules should not be allowed to take a test.

    d. Documentation regarding test taker rights also should be provided and explained.

17. If permitted by prevailing laws, test takers should be properly authenticated.[2] This may occur before, during or after a test. Acceptable methods of authentication include producing a government-issued photo ID, using biometric devices such as a fingerprint reader, palm reader or iris scanner, keystroke dynamics, or facial recognition.

18. During test administration and after authentication, tests are at greatest risk. For example, during this time, displayed items can be stolen and other forms of cheating may occur. In addition to prior agreed upon efforts during the planning and design stages, additional effort may be needed to ensure, to the extent possible, that test content cannot be stolen and that the probability of cheating is minimized[3].

    a. The test administration system should use a lockdown program or secure browser to restrict the operating system and testing workstation so that access to outside resources is limited to only those resources needed to complete a test.

    b. Proctors may have the ability to launch a test using special "keys" supplied by the test administration system. A similar key may also be provided to a test taker, so that both keys are required to launch a test.

    c. Proctors should monitor test takers vigilantly without being a distraction to them. If permitted by prevailing laws, proctoring can occur at a distance, online (through webcams), or locally (onsite and/or with CCTV).

    d. Proctors should have limited or no ability to view the test taker's screen or pages of a test booklet during a test.

    e. Proctors should be knowledgeable of expected methods for both cheating and test theft and well-trained in what to do if a security breach occurs, including the production of a test event irregularity report.

    f. Proctors should be sufficiently motivated to watch for security problems and to confront a test taker when a suspected breach occurs.

    g. Proctors should not have an interest or stake in a test's outcome. They should not be instructors or teachers for test takers nor familiar with the content covered by a test.

    h. If permitted by prevailing laws, cameras should be in place to assist in the monitoring, recording, and preservation of a test event and any security incidents.

[2] Authentication is not the same as identification. For high-stakes exams or assessments, it is not necessary to actually identify the person. It is only necessary to make sure that the person who will take the test is the same person who registered and signed up for the program.

[3] It is axiomatic that cheating can and will occur, even when maximally effective security measures are in place. It is the goal of a security program to minimize the effects of test theft and to reduce the incidents of cheating to manageable levels.

i. If possible and permitted by prevailing laws, devices capable of assisting a potential cheater or test thief (e.g., smart phones, tablets, cameras, papers) should be collected prior to the launch of a test and returned after the test.

j. The allowance of breaks should be managed carefully. After a break, test takers should not be able to review questions seen before the break.

k. Notepaper allowed, provided, and used during a test should be gathered afterward and dealt with according to program policies.

l. If cheating or capturing of test content is observed during a test, it should be dealt with swiftly and effectively, according to the specific guidance provided by the testing program. This may require the temporary suspension or permanent cancelling of a testing session, confiscation of equipment or materials used, and the completion of an official security irregularity report.

19. When digital test results are collected from remote servers, the data transfer should occur immediately after the completion of a test or after the completion of each item for online (e.g., Internet-administered) tests. The data should be protected by strong access procedures while residing on a remote server and strong encryption during transmission.

20. Tests and items should be evaluated regularly for indications of cheating or compromise. Item and test performance will change if items have been stolen and shared, and if cheating has occurred. Here are some examples:

    a. Abnormal response patterns (e.g., answering easy questions incorrectly and difficult questions correctly) may indicate cheating or theft.

    b. Abnormal response times for either tests or items (e.g. those that are abnormally short or long) may indicate a security breach or some other problem.

    c. Too many erasures on paper-based test answer forms, particularly wrong-to-right erasures, may indicate tampering or coaching.

    d. Unusual response similarity between pairs or groups of test takers may indicate collusion or proxy test taking behavior.

    e. Response and latency patterns that are similar for pairs or clusters of test takers may indicate collusion, proxy test taking, or coaching.

    f. Unusual gains from one testing session to another, whether for groups of test takers or for an individual, may indicate cheating.

    g. Unusual changes in item performance (e.g., in item statistical parameters) may indicate that an item has been compromised. Compromised items should be replaced immediately.

h. Differential item performance for one class of items on a test versus others may indicate the use of test content pre-knowledge. For example, the performance on Trojan Horse items (purposefully mis-keyed items) or embedded pre-test items (by definition have been exposed less) compared to the performance of operational, scored items may indicate the use of pre-knowledge.

i. If permitted by prevailing laws, test taker demographic data can be analyzed as well to determine possible fraud (e.g. proxy test taking behavior). For example, a test taker who lives in one country and has taken tests repeatedly in other countries in close time proximity may indicate collusive or proxy efforts.

j. Where test sessions have a regular schedule, test start and stop times can be monitored to make sure tests are given during regularly scheduled hours. Tests given outside of normal hours may indicate attempts to cheat or harvest content.

21. Software developed for exam authoring, exam delivery, or program management purposes must be developed using secure procedures that guard against common programming vulnerabilities and which is periodically evaluated (e.g., by third-party penetration tests).

22. Scoring technology-based tests generally occurs immediately after the test is completed and may occur during the test after each question has been answered (e.g., CAT). The threats and risks of cheating during the scoring process for such tests are minimal. For paper-based tests, the scoring process is lengthy and involves several steps, requiring more security measures to insure that scores are not altered.

    a. Scores may be provided provisionally and only confirmed after they have been subjected to a determination of their validity. This may include policies such that scores are not to be released or made official until irregularity reports have been reviewed and data forensics analyses have been completed and reviewed.

    b. Scoring of computerized tests should take place on well-protected remote servers not on the test taker's computer. For paper tests, when answer sheets are gathered and returned to a scanning or scoring location, tampering with the answer sheets can occur. A monitoring process should be in place to closely track and protect answer sheets in paper form until they can be processed for scoring purposes.

23. Tests, items, test results, and other important information (e.g., test taker demographic information), whether for paper-based or digital tests, often are stored for long periods (e.g. possibly years). Regardless of where or when these data are produced and gathered, professional procedures should be established to make sure that inappropriate access to this information (e.g., hacking) is extremely difficult, and that scores and other data cannot be accessed, modified, or deleted without proper authorization. ICT procedures and systems must be audited and updated periodically.

24. Before, during, and after a test administration window[4], a program should begin a process of monitoring the Internet for evidence of the disclosure of test content. Examples of this

[4] The period of time a test is available for administration.

disclosure may be individuals casually discussing a test or some of the questions, or it may be the accurate reproduction of one question or an entire set of test questions. When discovered, the program should send an appropriate request to the webmaster to stop the discussions, warn the participants, and remove the content. Stronger actions, including legal actions, should be considered if the material is not removed quickly.

25. Before, during and after a test administration window, a program should protect the content of a test from exposure to any individual who is not an authorized test taker or a representative of the testing program with rights to view the content.

Part 3: Responding to a Security Breach

A threat source occasionally is able to breach a program's defenses, resulting in successful cheating or stealing of test content. The following guidelines provide advice and support for dealing with these events. When it is discovered that cheating occurred, or that a test or items have been stolen, the testing program has a responsibility to investigate thoroughly, stop the breach, repair any damage, and take other appropriate actions. Actions should be taken to prevent the breach in the future.

The security committee should have full responsibility to respond to a security breach and be given authority to make decisions.

A program will learn of a possible or real breach in various ways, some better and easier to deal with than others. Here are a few:

- from a news reporter or other media
- from a proctor's irregularity report
- from a tip
- from data forensics reports
- from web monitoring reports
- from the automated security "systems" (e.g., use of inappropriate keystrokes during a test; attempted hacking)

Regardless of the source of a breach, the testing program needs to act quickly to determine the validity of the report and the extent of the breach. With this information, proper action can be taken. During a test, the monitoring or proctoring system should take immediate action when a breach is occurring. Before and after the test, the security committee is responsible for reviewing the details of the breach and responding accordingly.

Specific Guidelines on Responding to a Security Breach

1. A test taker's test should be paused or stopped if cheating or test theft is observed by proctors (e.g. either online or onsite proctors) or test administrators. An explanation should be provided to the test taker.

a. After questioning a test taker suspected of cheating or noncompliance with testing rules, a proctor or test administrator may allow the testing session to continue or decide that it should remain paused or be cancelled.

b. Any inappropriate material should be confiscated if permitted by prevailing laws. The test taker should be provided with reasons why this action is necessary.

c. The proctor, invigilator, or test administrator should complete a test irregularity report and forward it to the testing program security committee.

d. If the test is allowed to continue, a test irregularity report should be produced and reviewed by the security committee.

2. A breach should be investigated thoroughly to determine its pervasiveness and the extent of damage. Investigations may involve interviews with persons alleged to be involved or bystanders, data forensics analyses to see effects on scores, and/or web monitoring to check the range of test content disclosure.

3. A compromised test or set of items should be replaced as quickly as possible.

   a. Some testing programs may consider using a previously created test, also known as a "breach form," as a backup to replace a stolen test or set of items.

4. Scores shown to be inaccurate as a result of test fraud should be cancelled or invalidated.

   a. This process is simplified if policies are in place to routinely review scores and consider them as "provisional" and awaiting confirmation.

   b. If the invalid test scores have already been provided to test takers, then the test takers should be immediately contacted and informed that their scores are no longer valid, and that any decisions based on them will be reviewed.

5. Re-scoring a compromised test may be desirable if that enhances score accuracy for decision making (e.g., after the discovery and removal of compromised items).

6. Depending on testing program policies stated and on the amount of damage caused by a breach, additional action may be necessary, including civil or criminal legal action.

7. Websites that offer a program's copyrighted questions for sale without permission should be contacted with a demand that the content be removed from the website and all other locations. Several escalating steps are possible. The website should be closely monitored to verify that the content has been removed.

   a. Bystander letters from an agent of a program can be sent to the website informing the webmaster of the problems, and requesting that the material be removed. This has proved to be an effective first step with most website managers responding positively and quickly.

b. If the material remains on the site, more formal cease and desist letters can be sent that demand the site remove the content or face legal action. These notices can cite relevant country or regional statutes (e.g., USA's Digital Millennium Copyright Act or Europe's Copyright Directive).

8. Depending on policies and the severity of a breach, test takers may be requested and allowed to take the same test again using the same form or a different form. If permitted by prevailing laws, test takers involved in the cheating or test theft may be denied a re-test.

   a. Policy documents should state clearly the re-test rules as agreed to by all stakeholders, including test takers, before tests are administered.

   b. If available, a new test form should be used for re-testing.

   c. Additional conditions (e.g. a fee or a waiting period) may be applied as a condition for re-testing.

9. Following a breach, interest by stakeholders, third parties (e.g., media), and/or the public may be intense. Effective communication documents should be developed and distributed as quickly as possible. A public relations firm and or spokesperson may be useful. Materials prepared may include standardized communications announcing that a breach has occurred, the severity of that breach, and the actions taken to address that breach.

10. After a breach, a review of the existing security plan and associated procedures may be needed to decide if changes to existing security policies and procedures, including the addition of new ones, are warranted. These changes, if any, should be communicated to all stakeholders, and a new revision of the security plan should be prepared and approved.

TERMS AND DEFINITIONS

Authentication. The process of determining that the person sitting a test, or who sat a test, or who is preparing to sit a test, is the person who is supposed to take the test. Authentication is not the same process as identification, which attempts to actually identify the examinee.

Background Checks. The process of reviewing an individual's history in order to qualify him or her to help develop an exam.

Biometrics. Methods of collecting unique information about an examinee to be used for purposes of authentication or identification.

Breach. A successful attack by known or unknown threats.

Breach Test Form. An alternative form of a test which is used to replace a compromised test form.

Break. A rest time between sections of a lengthy exam.

Cheating. Any behavior that attempts to or succeeds in increasing a test score inappropriately.

Coaching. A form of collusion where one individual helps another answer test questions during the test.

Collusion. Individuals working together to either cheat on a test or steal the test content.

Compromise of Items and Tests. A determination that test content has been inappropriately exposed and may therefore no longer suitable for use in the test.

Data Forensics. Methods that analyzes the results of a test to detect patterns that might suggest cheating or test theft.

Differential Item Performance. A data forensics analysis of item performance under separate conditions (e.g., at test launch versus six months later) which might indicate that an item has been compromised.

Embedded Items. The process of placing non-scored items in a test intentionally to detect individuals cheating by using pre-knowledge.

Erasures. Responses on an answer sheet that have been erased.

Examinee. An individual sitting a test. Also referred to as a test taker.

Facial Recognition. A biometric method using webcam images where an examinee's facial features are compared at the time of program registration and just before test launch.

Forward-Only Item Presentation. Items are presented on the test without the possibility of returning to previously viewed items.

High-Stakes Assessments. Tests and other forms of assessments the results of which have significant consequences for an individual or organization.

Identification. The process of actually identifying the person sitting the test or preparing to sit the test. Identification is not the same process as authentication, which, in general does not attempt to identify the examinee.

Investigation. The process of determining the causes of and extent of a breach. Investigations may involve interviews, data forensics, analysis of procedures, review of reports, etc.

Invigilator. The individual responsible for the security of a test during test administration. Also called a proctor.

Irregularity Reports. Reports provided by proctors and others describing a cheating or other unusual factor that affected a test administration event.

Item Exposure. The exposure of an item during a test to individual examinees, or the exposure of the item after being illegally harvested and shared via the Internet or some other way.

Item Harvesting. Attempts, successful or not, to capture test content illegally and against program security rules.

Item Pools. Larger sets of items from which items are drawn to create a test either in advance of a test session or during the test session.

Keystroke Dynamics. A biometric method where an examinee's keyboard typing patterns are compared at the time of program registration and just before test launch.

Latency Analysis. A data forensics analysis of response latency, which is the response time from the moment the content of an item appears to an examinee to the point where the item has been answered and that answer has been submitted by the examinee. Unusually short or long latencies may indicate cheating or some other security problem.

Lockdown. A program launched just prior to the launch of an Internet-administered exam that restricts the examinee to keyboard and computer functions used solely to navigate and answer test questions. Access to other resources such as a computer's hard drive, the Internet, and certain key combinations are prohibited.

Proctor. The individual responsible for the security of a test during test administration. Also called an invigilator.

Provisional Scores. Unofficial scores provided to examinees after the completion of the exam, but which are subject to review by a security committee.

Proxy Test Taker. A person who takes tests for others.

Rescoring. The process of rescoring a test, perhaps after the influence of compromised items is removed.

Retakes. The retaking of an exam.

Re-testing. The process of allowing individuals to retake an exam.

Risk. An estimate of the likelihood of a threat becoming a successful breach, along with the amount of damage that a breach would cause.

Risk Analysis. An analysis of the various security threats to a testing program in order to estimate likelihood of risk, potential damage, and to allocate security resources appropriately.

Score Gains. A data forensics analysis of increases (or decreases) in scores to detect unusual changes that may indicate cheating.

Security Plan. A document describing the security policies and procedures of an organization.

Similarity of Responses. A data forensics comparison of the response patterns of two or more individuals in order to detect collusion, proxy test taking or coaching.

Tampering with Answer Sheets. A form of cheating where incorrect answers on an answer sheet are erased and replaced with correct answers.

Test Aids. Devices or documentation that may be used by an examinee while taking a test. Please note that the use of some test aids (e.g., calculators) may be allowed.

Test Taker. An individual sitting a test. Also referred to as an examinee.

Test Theft. Any behavior that attempts to capture or succeeds at capturing test content illegally.

Threat. An individual or method that has the potential to successfully cheat on an assessment or to successfully capture the test content.

Trojan Horse Items. Intentionally mis-keyed items embedded in an exam. The purpose of Trojan Horse Items is to detect an examinee using stolen items and answer keys to cheat on a test.

Verification Test. A test given at a later time to verify an examinee's performance on an earlier test.

Vulnerability. A weakness in the test security defense of a program.

Web Monitoring. A set of methods for searching the Internet for test questions from operational exams.

REFERENCES

Foster, D. F. & Miller, H. L., Jr. (2012). Global Test Security Issues and Ethical Challenges. In A. Ferrero, Y. Korkut, M. M. Leach, G. Lindsay, & M. J. Stevens (Eds.). *The Oxford Handbook of International Psychological Ethics* (pp. 216-232). Oxford: Oxford University Press.