Information Security Risk Management in Computer-Assisted Assessment Systems: First Step in Addressing Contextual Diversity

Hervé Cholez, Nicolas Mayer, and Thibaud Latour

Public Research Centre Henri Tudor

29, av. J. F.Kennedy, L-1855 Luxembourg, Kirchberg

{herve.cholez, nicolas.mayer, thibaud.latour}@tudor.lu

Abstract

Security in Computer-Assisted Assessment (CAA) systems is a complex and context-dependent problem. The large diversity of contexts, stakes, and processes makes it hard to depict the full situation of CAA security. Hence, identifying all risks with a systematic method that takes into account contextual variability is essential to providing adequate solutions and to treating these risks. This paper presents a first contribution to the assessment of security risks related to CAA systems by using security risk management. In this paper, our approach consists of identifying the security needs of CAA systems and developing an overview of CAA security needs derived from interviews with four experts. The salient parts are highlighted and different groups with similar needs are identified. Our results are an essential step towards performing a full risk assessment and lead to a better understanding of the security issues in many CAA types and contexts.

One of the main concerns in CAA (Computer-Assisted Assessment), and particularly in summative high-stake assessment, is security (Rabinowitz & Brandt, 2001; Marais *et al.*, 2006; etc). Security is particularly challenging due to the wide range of attacks these systems have to handle. As for any IT system, "classical" technical attacks can occur within the Information System (IS). However, particular forms of attack pertaining to the specific processes and stakes of CAA can also be identified,

including the "braindump" of test items (Smith, 2004). The scope of these domain-dependent security risks strongly varies depending on the assessment needs, which depend on a series of contextual parameters depicting the types of assessment.

Currently, no exhaustive and contextualised database of CAA systems security exists. It is thus difficult to define a standard secure CAA system adequately protected against all relevant risks and aligned with the sensitivity of the system. Considering the high importance of security considerations and the current lack of systematic security risk assessment, such a database is needed. Our research focuses on defining the different kinds of security risks targeting CAA systems and, in this paper, the objective is to define the security needs of CAA systems. Attacks can occur in the context of summative assessment (e.g., test for language certification), formative assessment (e.g., a self-assessment for evaluating weaknesses in a given subject matter); either in high-stake situations (e.g., an exam for obtaining a diploma) or low-stake situations (e.g., preparation for an exam); on a large scale (e.g., a national assessment) or a low scale (e.g., a single classroom test). These attacks are not all the same and have different impacts on the system; thus, we need to consider all types of assessments that use CAA systems.

Section 1 of this paper provides the security risk management background for our research work. Section 2 presents the state of the art relative to CAA security. Section 3 presents our research methodology for identifying and characterising all types of security risks targeting CAA systems. Finally, Section 4 discusses our findings and future work.

1. Security Risk Management Process

The objective of Information System Security Risk Management (ISSRM) is to protect an organisation's assets from all accidental or deliberate harm to IS security; ISSRM achieves this by using a risk management approach. ISSRM activities usually follow an overall process composed of classical steps generally found in traditional ISSRM methods (e.g., EBIOS (EBIOS, 2004), ISO/IEC 27005 (ISO 27005, 2008), etc.). Below is an overview of the steps composing the overall process (Mayer, 2009):

Context and asset identification: The process starts with a study of the organisation's context and the identification of its assets. An asset is anything that has value for the organisation and that must be protected. In this step, the organisation and its environment are described, focussing on the sensitive activities related to information security. If an Information System is already in place, an overview is created.

Determination of security objectives: The security needs of the organisation are defined next. Based on the asset identification, the security objectives to be achieved need to be determined for each asset. Security objectives are often defined in terms of the assets' confidentiality, integrity, and availability properties.

Risk analysis and assessment: The main step of the process is the risk analysis, which elicits those risks that harm the assets and threaten the security objectives. This step consists of identifying risks and estimating their level in a qualitative or quantitative manner. We deal with risk assessment (ISO/IEC 27005) only when the level of analysed risks has been evaluated against the security needs determined during the second step of the process.

Risk treatment: Once risk assessment has been performed, decisions about risk treatment are made. Risk treatment measures can include avoiding, reducing, transferring, or retaining risk (ISO/IEC 27005).

Security requirements definition: Security requirements can now be determined as security solutions for mitigating the risks, mainly if the risk reduction treatment was chosen. However, security requirements can emerge from other treatments, such as risk transfer, which generally calls for some third-party requirements. The selection of security requirements is principally based on the ratio between the cost of the treatment and the estimated benefit.

Control selection and implementation: Requirements are finally instantiated into security controls, *i.e.*, system-specific countermeasures implemented within the organisation.

2. CAA Security

Overview of the Security Issues in CAA

In recent years, research has focussed on studying different security issues. For instance, test-takers use computers in CAA, therefore if they want to cheat, they have access to a lot of help, such as calculators, Internet, spell checker, etc. A test-taker can even use spyware or sniffers to gain access to the answers of another student (Rowe, 2004). Laubscher *et al.* (2005) studied the detection of these particular cheating methods and Kinnersley *et al.* (2001) proposed prevention solutions. For the purpose of authentication and to prevent cheating, Ko & Cheng (2004) proposed using random video monitoring to capture test-takers' faces at random intervals during tests, and Jung & Yeom (2009) suggested establishing "full video" monitoring, including audio recording. One problem specific to high-stake assessments is braindump, especially when the target population does not take the test simultaneously, which is often difficult to ensure in large-scale assessments (Rowe, 2004). Some examinees memorize ("brain") test items and share ("dump") the information after the test. Many specialised websites disseminate test items, while some community websites are dedicated to test-taker preparation but do not pay too much attention to users' posts (Smith, 2004) and legal issues regarding intellectual property rights. To eliminate this problem, the creation of larger item banks has been proposed (Rabinowitz & Brandt, 2001). However, constructing high-quality questions is difficult, time-consuming, and expensive (Sim *et al.*, 2004), since such banks usually require thousands of questions (Maughan *et al.*, 2001). Another way to prevent braindump and cheating is to concentrate efforts on item design. Dynamic questions (McGough *et al.*, 2001) or "*Multiple choices with new media distractors*" in Scalise & Gifford's taxonomy (2006) enable the generation of a "new" question for each test-taker with the same level of difficulty (e.g., variables in the question as random integers). Though this solution looks appealing, it remains very specific and problematic in terms of psychometric validity and comparability.

Other research has focussed on one particular domain or one specific security criterion such as confidentiality, integrity or availability. In order to improve confidentiality with identity management and test-taker authentication, Barker & Lee (2007) compared different biometrics authentication systems such as fingerprint and retinal eye-pattern recognition. Kinnersley *et al.* (2001) proposed using "one-time passwords" for authentication. In addition, Bartram (2006) proposed adding IP checking during the authentication procedure, even if it is not sufficient to prevent imposture in a cheating purpose. Test delivery can be restricted in order to increase

security. Test integrity for a low-scale test, for example, can be ensured by making it available on the server only a few hours before the examination starts (Aojula *et al.*, 2006). However, a server may fail and affect the availability of the test. To prevent this, a back-up server can operate throughout the assessment and a full set of printed examination papers can be available in a sealed envelope (Aojula *et al.*, 2006). Of course, the paper-and-pencil backup might not be psychometrically equivalent to the computerised counterpart and restricts the benefits of CAA.

Few works addressing the whole security issues of CAA can be found in the literature. For instance, Marais *et al.* (2006) present security issues specific to CAA. However, his review does not cover very specific aspects (e.g., braindump is not included) and is not adapted to the full range of assessment types and contexts. In terms of standardisation, some work has been done recently around the ISO/IEC 23988 standard (ISO 23988, 2007). This standard contains a two-page section dedicated to CAA security and proposes some global good practices (e.g., "*security of items and correct responses*"). However, it does not concentrate on specific issues but more on very general issues and does not take into account the specificities of different assessment contexts (e.g., low-stake formative CAA does not require the same security countermeasures as high-stake summative CAA).

3. Security Needs in CAA

As depicted in Section 2, security in CAA has been addressed from a very narrow viewpoint, concentrating on particular aspects pertaining to the CAA situations of interest. As a consequence, no global approach to security in CAA systems has been proposed so far. Indeed, while many authors have described research about dedicated security issues or dedicated security criteria, little of that can be transferred to the different types of CAA and most were restricted to one particular CAA domain (e.g., high-stake summative). In addition, for each CAA security aspect, different studies established their own isolated countermeasures. Adopting a different and more global stance, our research objective aims at addressing the full scope of assessment types and defining the security needs and specifics risks associated with CAA in a more integrative vision. As a direct benefit, establishing a consolidated database with these results will allow organisations to quickly define the appropriate security requirements for their CAA system.

Research Method

In order to achieve the goal of establishing a consolidated CAA security database, we apply an ISSRM approach to CAA systems, as presented in Section 1. In this paper, we shall focus on the first two steps of the security risk management process, *i.e.*, *Context and asset identification* and *Determination of security objectives*. The research method used to achieve this objective can be divided into three steps:

Step 1: CAA assets identification. In the first step, we identify the CAA assets. Although assessment procedures may vary, we use the typical assessment lifecycle from the international standard ISO/IEC 23988 (ISO 23988, 2007). For each process, different kinds of information with different security needs may be used. For example, if an organisation uses sensitive information (e.g., religious information), its security needs are definitely not the same as for "classical" personal information (e.g., age, gender). These different pieces of information are the CAA assets. Hence, for each process defined by ISO/IEC 23988, different pieces of information will be identified. This constitutes our process and information reference model.

Step 2: Determination of the scale of security needs. A scale of security needs will be defined in this second step. This scale will be used to determine the security needs of each asset identified in Step 1. The scale consists of different and distinct levels for the three classical security criteria: confidentiality, integrity, and availability (ISO 27005, 2008). This scale will be the reference used with each interviewee to define the level of security needs relative to the corresponding asset. Since the final result will be averaged over several experts, this scale must be consistent, coherent, understandable by every interviewee and specific to CAA systems.

Step 3: Determination of security needs. Finally, we shall define the security needs for each type of CAA system. Experts' interviews will be used to establish a clear overview of the different security needs in each assessment type. For each piece of information of each assessment type, experts will evaluate the security needs level based on the scale defined in Step 2. Interviews allow obtaining comments from experts and avoid misunderstandings. The average over all experts will be calculated and used as a final value.

Establishing the Security Needs in CAA

The first step of our research method was carried out to establish our process and information reference model. Based on the ISO/IEC 23988, the list of processes of the typical lifecycle of an assessment was identified and organised in the left column of our reference model, as illustrated in Table 1. The focus of this work is not on preparing assessment content, because in the vast majority of situations, these processes, which involve no interaction between organisations and test-takers, are internal to an organisation and subject to its general IT security policy (e.g., access rights, server backup, etc.). Following this scope, "*Identification of need to assess*", "*Design of outcomes/assessment methodology*", and "*Preparation and calibration*" are excluded here. We identified 14 pieces of information related to 10 processes, as illustrated in Table 1 with some examples. The security needs relative to each of these will be evaluated in Step 3 using the scale defined in Step 2. The completeness and relevance of each piece of information was evaluated, validated and completed by each interviewee during Step 3. We noticed that no information is linked to the "*distribution*" process, since all information used by this process is attached to other processes and the distribution process only provides this particular information.

The aim of Step 2 is to establish a scale of security needs depicting the level of importance associated with each piece of information attached to the key processes of the reference model. After several tests of the scale, the final levels of needs collected from experts and described in Table 2 are on a 6-level scale ranging from 0 to 5 for each security criterion, *i.e.*, confidentiality, integrity, and availability. Increasing numbers denote increasing importance of the security needs. A value of zero means that there is no security requirement, e.g., a confidentiality level of 0 implies that the information can be accessed by everybody and that there is no need to restrict its access. At the opposite side, a confidentiality level of 5 implies that it is absolutely essential that nobody can access the information. Intermediate values correspond to intermediate stakes as defined in Table 2. In order to maintain consistency among the experts and ensure good validity of the collected data, it is essential that the definition of each level is clear, unambiguous, and distinctive.

The third step consists of the quantitative evaluation of security needs by all interviewees. This evaluation is done for each criterion: Confidentiality, Integrity, and Availability (named respectively C, I, and A in Table 3) by using the scale defined in Step 2. However, as already mentioned above, security needs may strongly depend

on the type or context of the assessment. In order to reflect this dependency, the evaluation of security needs was done for all different types of CAA systems. At the end, four CAA experts with different profiles (e.g., security, development, management, etc.) were individually interviewed and the mean of all results was calculated. The final results are shown in Table 3 and discussed below.

Table 1. Processes and related information reference model in a classical assessment lifecycle, from the ISO/IEC 23988 standard

Processes	Information	Examples
Identification of need to assess	*Out of scope*	*Out of scope*
Design of outcomes / assessment methodology	*Out of scope*	*Out of scope*
Preparation and calibration	*Out of scope*	*Out of scope*
Pre-registration (includes payment)	Personal information	Address, age, email, background information (diploma…), etc.
	Payment information	Credit Card number, CVV number, etc.
	Sensitive information	Ethnics, political information, etc. (specific personal information with different security level requirements)
Distribution	*N.A.*	*N.A.*
Authentication (includes identification)	Authentication information	Password, Photography, biometrics (fingerprint, palm vein scan…), etc.
Delivery	Test questions	All questions/items of the test, including stems, alternatives, medias, etc.
Response return	Testees answers	Answers selected, draw, write, etc. by test-takers
	Logs	Time spend on a question, all different answers selection before validation by test-takers, mouse movement, etc.
Scoring, result determination and/or feedback	Correct answers	Correct/expected answers of each questions
	Scoring rules	Scoring algorithm and parameters
Data return	Results (score)	Marks, fail/pass, etc.
Analysis	Feedback	Analysis of results, personal feedback, etc.
Appeals	Appeal information	Relevant evidences for forensic send by test-takers including personal information, testee answers, etc.
	Result appeal	New Results and/or Feedback with explanation
Certification	Certification Information	Diploma, certificate, etc.

Discussion

First, we can observe that the summative, large-scale, and high-stake CAA is the most sensitive form of CAA, since the security needs are significantly higher than for other types of assessments. Formative, low-scale and low-stake CAA is the least sensitive form. This finding obviously corresponds to the intuitive notions attached to these assessment contexts.

Second, we can see from Table 3 that the needs in large-scale and low-scale assessments are very similar. This can be explained by the fact that the security needs for an assessment distributed in a large-scale and in a low-scale environment are also similar, even if the risks in a large-scale assessment are stronger and more numerous. For example, the test questions for a summative high-stake test should be kept absolutely confidential and the access restricted to the test assessors: this corresponds to level 4 of the security needs, independent of the assessment scale. This justifies the same confidentiality level of 4 for both large- and low-scale

assessments, even if the large-scale context is potentially exposed to more risks (e.g., braindump) than the low-scale one. The risk assessment step will be done on the next step of the Risk Management and is thus part of our future work.

Table 2. Scale for the evaluation of security needs

	Confidentiality	Integrity	Availability
0	No confidentiality needs	No integrity needs	No availability needs
1	Low confidentiality needs (access to all test participants)	Low integrity needs (assurance to have information complete and correct occasionally)	Low availability needs (information available in 3-7 days)
2	Medium confidentiality needs (access to a group)	Medium integrity needs (assurance to have information complete and correct in the day)	Medium availability needs (information available in 1-3 days)
3	Substential confidentiality needs (personal access + test assessors access)	Substential integrity needs (assurance to have information complete and correct after the test/the processing)	Substential availability needs (information available after some hours)
4	High confidentiality needs (test assessors access)	High integrity needs (assurance to have information complete and correct after each modification)	High availability needs (information available between 0 and 2 hours)
5	Essential confidentiality needs (can't be accessed)	Essential integrity needs (assurance to have information complete and correct at any moment)	Essential availability needs (information available at any moment)

Table 3. Globally averaged security needs by CAA types

	Formative												Summative											
	Large-Scale						Low-Scale						Large-Scale						Low-Scale					
	L-stake			H-stake			L-stake			H-stake			L-stake			H-stake			L-stake			H-stake		
Information	C	I	A	C	I	A	C	I	A	C	I	A	C	I	A	C	I	A	C	I	A	C	I	A
Personal information	3	2	2	3	4	2	3	2	1	3	4	2	3	4	2	3	4	3	3	4	2	3	4	3
Payment information	5	5	3	5	5	4	5	5	3	5	5	4	5	5	3	5	5	4	5	5	4	5	5	4
Sensitive information	5	2	1	5	4	2	5	2	1	5	4	2	5	4	2	5	5	3	5	4	2	5	5	3
Authentication info.	5	3	3	5	4	4	5	3	3	5	4	4	5	4	3	5	5	5	5	4	3	5	5	5
Test questions	1	2	2	3	4	4	1	2	2	3	4	4	4	4	3	4	5	4	4	4	3	4	5	4
Testees answers	3	2	2	3	4	4	3	2	2	3	4	4	3	4	3	4	5	4	3	4	3	4	5	4
Logs	4	2	1	4	4	3	4	2	1	4	4	3	4	3	2	4	5	3	4	3	1	4	5	3
Correct answers	1	2	2	4	4	3	1	2	2	4	4	3	4	3	3	4	5	4	4	3	3	4	5	4
Scoring rules	1	1	1	4	4	3	1	1	1	4	4	3	4	4	3	4	5	4	3	3	3	4	5	4
Results	3	2	2	3	4	3	3	1	2	3	4	3	3	3	2	3	5	3	3	3	2	3	5	3
Feedback	3	2	1	3	4	2	3	1	1	3	4	2	3	3	2	3	5	2	3	3	2	3	4	2
Appeal information	3	2	1	3	3	3	3	2	1	3	3	3	3	3	1	3	5	3	3	3	2	3	5	3
Result appeal	3	2	1	3	3	2	3	2	1	3	3	2	3	3	1	3	5	2	3	3	1	3	5	2
Certification Info.													2	4	1	3	5	2	2	4	1	3	5	2
Averages	*3,0*	*2,2*	*1,7*	*3,6*	*3,9*	*3,0*	*3,0*	*2,1*	*1,6*	*3,6*	*3,9*	*3,0*	*3,5*	*3,7*	*2,3*	*3,9*	*4,7*	*3,3*	*3,5*	*3,6*	*2,3*	*3,9*	*4,7*	*3,2*
	2,30			*3,51*			*2,24*			*3,49*			*3,18*			*3,95*			*3,15*			*3,93*		
	2,90						*2,87*						*3,57*						*3,54*					
	2,88												*3,55*											

Finally, by analyzing Table 3, we can define some groups of information that need to be protected in the same way. The confidentiality of "Personal information", "Payment information", "Sensitive information", and "Authentication information" has the same weight in all assessment contexts. This means that the confidentiality of

these four assets is context-independent and that in all types of CAA, these assets should have the same level of protection. However, the integrity and availability of these assets can be slightly different, such as the authentication process in low-stake CAA that can be interrupted for a longer period than in high-stake CAA.

"Personal information", "testee answers", "results", "feedbacks", "appeal information", and "result appeal" were attributed equivalent security needs. All of these pieces of information are related to data used to compute the scores and are bound to the test-taker, making them equivalent to personal data. This group of information shares approximately the same level of needs in terms of confidentiality, integrity, and availability. However, depending on the assessment type, some adjustments can be introduced.

In the same manner, "test questions", "logs", "correct answers", and "scoring rules" are quite equivalent in terms of security needs. They refer to the test itself and are managed by test assessors.

4. Conclusion and Future Work

The objective of this research work is to establish the different kinds of security risks targeting CAA systems and in this paper to define the security needs. Firstly, the security risk management process was introduced. Then the state of the art of security in CAA systems was outlined. Finally, the first two steps of the CAA security risk management were performed. As a conclusion, the security needs of each type of CAA systems were established, a sensitive area in CAA systems was identified (summative, high-stake, large-scale CAA), and different information groups were defined based on their sensitivity.

The relevance of our results has to be confronted to the particularities of our research work. This work was done at our institution with CAA experts working on the TAO project. TAO (French acronym for "computer-assisted testing") is an open and versatile CAA platform (Plichart *et al.*, 2004); designed for all types and contexts of CAA. Consequently, experts with knowledge and strong experience in all types of CAA were available for interviews. However, we do not want to remain restricted to this community. Therefore, the next steps of our work will be to consolidate these results with more CAA experts' interviews in different organisations, countries, etc, while further refining the context definition with a wider set of variables.

As part of our future work, the next risk management steps will be performed, mostly defining the threats and vulnerabilities impacting CAA systems and defining related security requirements.

References

Aojula, H.; Barber, J.; Cullen, R. & Andrews, J. (2006). Computer-based, online summative assessment in undergraduate pharmacy teaching: The Manchester experience. *Pharmacy Education*, 6(4), 229-236.

Barker, T. & Lee, S. (2007). The verification of identity in online assessment: A comparison of methods. *Proceedings of 11th Computer Aided Assessment Conference*.

Bartram, D. (2006). The Internationalization of Testing and New Models of Test Delivery on the Internet. *International Journal of Testing*, 6(2), 121-131.

EBIOS (2004). Expression of Needs and Identification of Security Objectives. http://www.ssi.gouv.fr/en/confidence/ebiospresentation.html.

ISO/IEC 23988 (2007). *Information technology - A code of practice for the use of information technology (IT) in the delivery of assessments*. International Organization for Standardization.

ISO/IEC 27005 (2008). *Information technology - Security techniques - Information security risk management*. International Organization for Standardization.

Jung, I.Y. & Yeom, H.Y. (2009). Enhanced security for online exams using group cryptography. *IEEE Transactions in Education*, 52(3), 340-349.

Kinnersley, N; Mayhew, S. & Hinton, H.S. (2001). The design of a web-based computer proficiency examination. *Proceedings of 31st Annual Frontiers in Education Conference - Impact on Engineering and Science Education*.

Ko, C.C. & Cheng, C.D. (2004). Secure Internet examination system based on video monitoring. *Internet research: Electronic networking applications and policy.*

Laubscher, R.; Olivier, M.S.; Venter, H.S.; Rabe, D.J. & Eloff, J.H.P. (2005). Computer Forensics for Computer-Based Assessment: The Preparation Phase. *Proceedings of 5th Annual Information Security South Africa Conference.*

Marais, E.; Argles, D. & von Solms, B. (2006). Security Issues Specific to e-Assessments. *Proceedings of 8th Annual Conference on WWW Applications.*

Plichart, P.; Jadoul, R.; Vandenabeele, L. & Latour, T. (2004). TAO, A Collective Distributed Computer-Based Assessment Framework Built on Semantic Web Standards. *Proceedings of the International Conference on Advances in Intelligent Systems - Theory and Application AISTA2004.*

Maughan, S.; Peet, D. & Willmot, A. (2001). On-line formative assessment item banking and learning support. *Proceedings of 5th International Computer Assisted Assessment Conference.*

Mayer, N. (2009). *Model-based Management of Information System Security Risk*. PhD Thesis.

McGough, J.; Mortensen, J.; Johnson, J. & Fadali, S. (2001). A web-based testing system with dynamic question generation. *Proceedings of 31st Annual Frontiers in Education Conference - Impact on Engineering and Science Education*.

Rabinowitz, S. & Brandt, T. (2001). Computer-Based Assessment - can it deliver on its promise? *WestEd research report produced under contract to the U.S. Department of Education*.

Rowe, N.C. (2004). Cheating in Online Student Assessment: Beyond Plagiarism. *Online Journal of Distance Learning Administration*, 7(2).

Scalise, K. & Gifford, B. (2006). Computer-Based Assessment in E-Learning: A Framework for Constructing "Intermediate Constraint" Questions and Tasks for Technology Platforms. *Journal of Technology, Learning and Assessment*, 4(6).

Sim, G.; Holifield, P. & Brown, M. (2004). Implementation of computer assisted assessment: lessons from the literature. *ALT-J Research in Learning Technology*.

Smith, R.W. (2004). The Impact of Braindump Sites on Item Exposure and Item Parameter Drift. *Proceedings of Annual Meeting of the American Education Research Association.*