SCIENCE & SECURITY July 2014

CASRA NEWSLETTER – ISSUE 7

CASRA is well-known for its applied research and the solutions for the needs and challenges of project partners and customers. We are happy to present you two articles that exemplify this by showing user-oriented approaches towards practical problems.

In the section "Research Put Across", we report on an applied research project in Doha, Qatar, that has been successfully established on a long-term basis to provide our project partners with results on the competency level and development of their airport security officers.

In the section "Security in Practice", we give you an insight into how we identified the practical needs and requirements regarding X-ray image interpretation and knowledge testing to develop a flexible software tool that is suitable to cover many different testing scenarios.

With this issue of "Science & Security" we wish you a lovely summer!

Prof. Dr. Adrian Schwaninger

Dr. Diana Hardmeier

TOPICS IN THIS ISSUE:

RESEARCH PUT ACROSS

THE DOHA INTERNATIONAL AIRPORT AND HAMAD INTERNATIONAL AIRPORT PROJECT

CASRA's main aim is to increase security and facilitation at airports and in other environments involving people and technology. In the past decade, CASRA researchers have been able to conduct numerous applied research projects with partners who share this aim, many of them on a long-term basis. An example for a successful long-term collaboration is the research and development project conducted by CASRA scientists in Doha together with the Civil Aviation Authority of the State of Qatar.

SECURITY IN PRACTICE

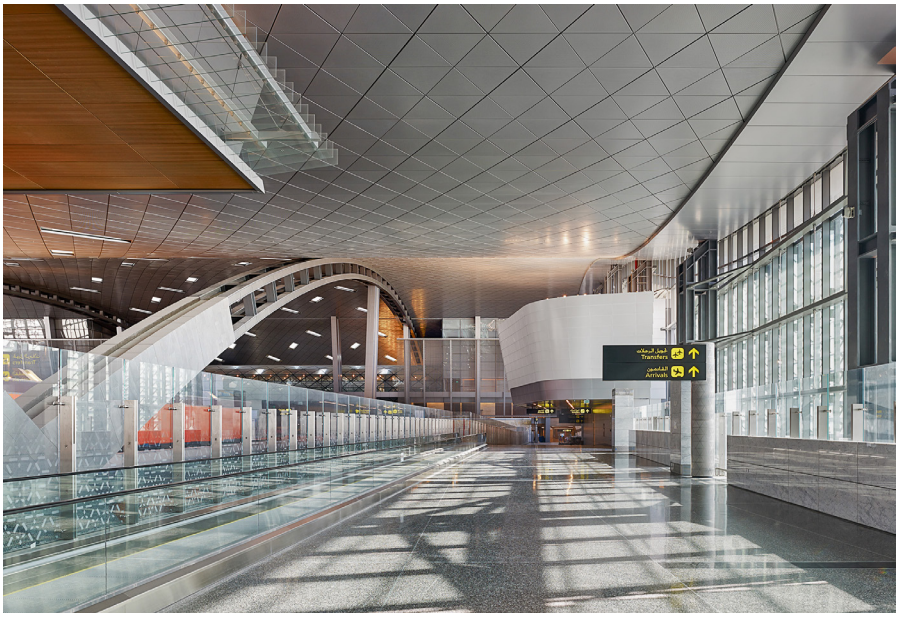
SECURITY IN PRACTICE USING THE CASRA TESTING PLATFORM (CTP)

Tests are common in every area of life: They are necessary whenever we have to prove that we possess certain abilities, that we have learned something or that we are able to put newly acquired skills to practical use. Based on scientific know-how and several projects with regulators, airports and security companies, the CASRA Testing Platform (CTP) was developed. It allows testing X-ray image interpretation competencies and the theoretical knowledge of screeners in an effective, efficient and flexible way.

THE DOHA INTERNATIONAL AIRPORT AND HAMAD INTERNATIONAL AIRPORT PROJECT

Text: Sandrina Ritzmann & Stefan Michel

CASRA's main aim is to increase security and facilitation at airports and in other environments involving people and technology. In the past decade, CASRA researchers have been able to conduct numerous applied research projects with partners who share this aim, many of them on a long-term basis. An example for a successful long-term collaboration is the research and development project conducted by CASRA scientists in Doha together with the Civil Aviation Authority of the State of Qatar.



The project's main aim has been to maintain and further increase the high security standard at Doha's international airports. Launched in 2008 at Doha International Airport the project is now continued at Hamad International Airport, the new international Qatari airport which opened in April 2014. One of the main success factors of this project has been the close collaboration between CASRA and the end users in Qatar to plan and implement the X-ray training and testing measures. In the remainder of this article, the scope of the project and a number of insights from the research results are shared.

SCOPE OF THE PROJECT

Previous studies in the areas of cabin baggage screening and hold baggage screening showed that computer-based training (CBT) of X-ray image interpretation competency effectively improves threat detection performance (e.g. [1], [2]). Based on these results, the X-Ray Tutor (XRT) was introduced as a training tool for airport security officers (screeners) in these two areas. It was an exigency to provide the XRT in Arabic, which CASRA was ready to comply with.

The objective of training with the XRT is to enhance the visual knowledge base and consequently the X-ray image interpretation competency of screeners by showing them a large number of different threat items in different bags. This targets the so-called knowledge-based factor influencing threat detection ([3]). Four threat item categories are used: Guns, knives, Improvised Explosive Devices (IEDs) and other prohibited items. Depending on the training area (cabin or hold baggage screening), either all four categories or only IEDs are employed (e.g. for training of hold baggage screeners). To ensure a maximum training effect, the majority of IEDs used for all CASRA solutions are built with real explosives. All training images are merged "on-the-fly" by the software based on two image libraries, one containing threat items and the other containing bags without any threat items. Image difficulty is individually adapted to the performance level of the screener based on the three image-based factors viewpoint (is the threat object rotated or not), superposition (to what extent is the threat object superimposed by other objects) and bag complexity (how many different items are in the bag). Previous research results indicate that these characteristics of X-ray images influence threat detection substantially ([4]).

To continuously measure and monitor the effects of CBT with the XRT, regular tests using the X-Ray Competency Assessment Test (X-Ray CAT) have been conducted to inform the project partners of the level of image interpretation competency of their staff. For the area of cabin baggage screening, each test is built of 256 images, half of them featuring a prohibited item from one of the four categories guns, knives, IEDs and other prohibited items. For the area of hold baggage screening, 160 images are used, half of them featuring IEDs as prohibited items. The tests fulfill high scientific standards and are reliable, valid and standard-

ized ([5]). Similar tests are also in use for the legally required airport security screener certification in several European countries. From an organizational point of view, the X-Ray CAT can serve to check and adjust training schedules and modalities if, for example, performance does not increase as expected (e.g. introducing specific IED training or requiring more training sessions per month). The results the project partners are provided with after analysis of the test data can also be used to give the screeners feedback on their detection performance (effectiveness) and on their average time needed to analyze X-ray images (efficiency). This can be a helpful addition to the inbuilt performance measurements in XRT, because the X-Ray CAT is standardized and the same for all test-takers (as opposed to the individually adaptive XRT training sessions).

In addition to CBT and regular performance measurements, the X-Ray Object Recognition Test (X-Ray ORT) was introduced as part of the project. The X-Ray ORT measures the ability of a person to inspect and analyze X-ray images in order to detect threats ([6]). Similar to ability differences when people learn a new language, not everybody is equally able to become a good screener. More specifically, not everybody is equally able to deal with the image-based factors described above that determine image difficulty. The X-Ray ORT consists of 256 grey-scale images. Half of them contain a prohibited item, more specifically a gun or a knife. When the test was constructed, the use of color images and other prohibited items was abandoned to make the test suitable for selection purposes. In pre-employment assessments, novices without any experience interpreting X-ray images should be able to take the test. In the Doha project, the X-Ray ORT was used to get an overview of the competencies of the existing staff. In the future, it is foreseen to implement the test in the selection process.

INSIGHTS FROM RESEARCH RESULTS

At the start of the project in 2008, a baseline had to be established to analyze changes in the X-ray image interpretation competency achieved through training. Therefore, the screeners took the X-Ray CAT prior to starting CBT with the XRT. In the following, the outcomes from the tests in the area of cabin baggage screening are shown. The results from 2008 are compared to the results of another X-Ray CAT that was conducted by all screeners at the end of 2013. The results of both measurements are displayed in Figure 1 using the detection performance measure A'. A' takes into account the hit rate (how many threat items were correctly identified) as well as the false alarm rate (how many harmless bags were wrongly judged as not ok). Due to security reasons, the exact levels of the detection performance measure A' are not displayed.

Figure 1 shows that the detection performance for guns and knives was already quite high when the project started. This can be explained by the fact that the visual appearance of guns and knives is common knowledge even for people without any security screening experience. Most people have a more or less extensive mental "visual library" covering different shapes and forms of guns and knives (knowledge-based factor of image interpretation, see above). This knowledge can, to some extent, be used to interpret X-ray images. Training at Doha International Airport further enhanced the visual knowledge of what different types of guns and knives look like in X-ray images, apparent in the increase in detection performance.

Inspecting the other results in Figure 1, the large increase in detection performance for IEDs is very pronounced. Again, the reason lies in the visual knowledge of screeners. IEDs are hardly encountered at work or in everyday life and can take many different appearances. Training with a large number of different IEDs enhances the visual knowledge base and highly increases the chance that similar objects

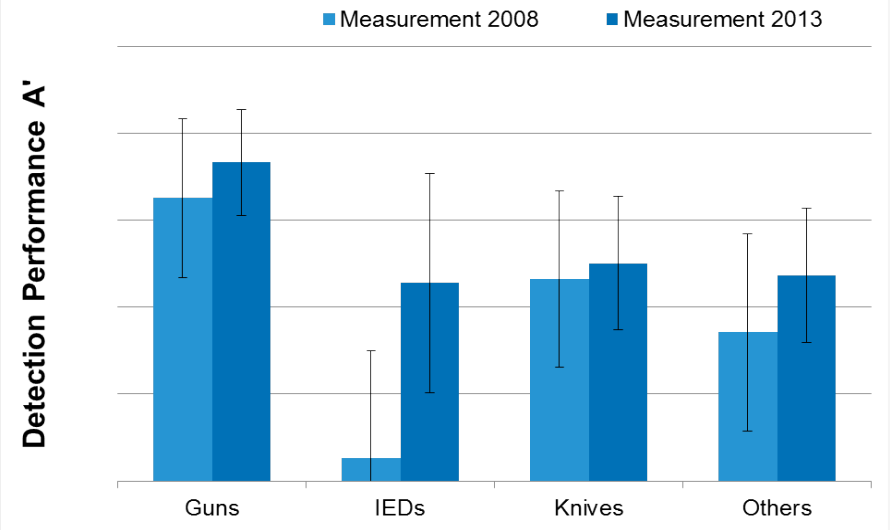


Figure 1: Mean detection performance A' with standard deviations for the two measurements 2008 and 2013.

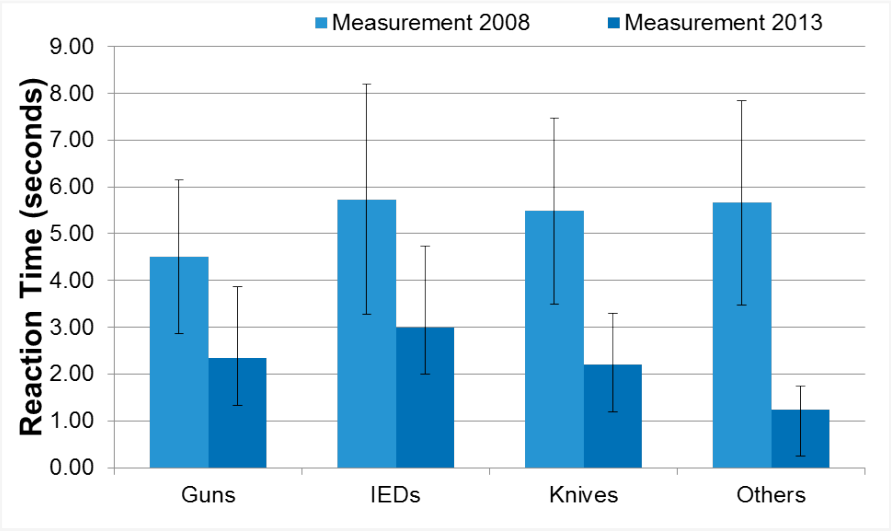


Figure 2: Mean reaction time (seconds) with standard deviations for the two measurements 2008 and 2013.

are recognized at the work place or in a test. This result clearly shows the importance and also the efficiency of regular CBT, which allows for a large number of processed images per training unit. The more objects a screener has stored in his visual memory, the larger the "database" that can trigger recognition of a prohibited item, even if the specific object in question was never seen before. This is the so-called transfer effect of training ([2]) which allows recognizing new, but similar looking objects.

A significant increase could also be observed for the category of other threat items. This category contains various objects ranging from electric shock devices to throwing stars or fireworks. Here again, the larger the number of different objects that are trained, the higher is the likelihood that a threat item is detected.

In addition to the effects on detection performance, our data showed that the reaction times for all threat categories significantly decreased, indicating that as a result of training, more accurate decisions are made in a shorter amount of time (see Figure 2). Consequently, well-trained screeners are not only crucial for security, but potentially also for the throughput achieved at a checkpoint.

Overall, CBT with the XRT at Doha International Airport effectively increased threat detection performance, which is in line with previous studies ([1], [2]).

OUTLOOK

Aviation security at an airport is a topic that needs continuous attention, not only because of the ever-changing threat situation, but also because of the human factors that are so important in the security system: Training and competency assessments as well as personnel selection procedures, aiming at improving the X-ray image interpretation competency of the screener workforce, can only unfold their full impact over a certain period of time.

Therefore, the collaboration with the Civil Aviation Authority of Qatar will continue in the future to maintain and further improve the detection performance of screeners at Hamad International Airport. Training will be accompanied by continuous performance measurements using the X-Ray CAT. As mentioned above, the X-Ray ORT will be introduced as a pre-employment assessment tool to assure that only people with the necessary aptitudes and abilities to become a profes-

Figure 3: Screenshot of XRT3 in Arabic

sional and skilled screener get hired for the job.

Regarding the Arabic version of the XRT, an upgrade to the newest release 3.7 (see Figure 3) is previewed for the second half of 2014, together with the installation of additional training workstations at the new airport.

Further information on the CASRA solutions employed in this project can be found in previous issues of our Newsletter: Newsletter No. 1 (X-Ray ORT, X-Ray CAT) and Newsletter No. 2 (X-Ray Tutor). Please request them from info@casra.ch.

REFERENCES

[1] Michel, S., de Ruiter, J.C., Hogervorst, M., Koller, S.M., Moerland, R., & Schwaninger, A., (2007). Computer-based training increases efficiency in X-ray image interpretation by aviation security screeners. *Proceedings of the 41st Carnahan Conference on Security Technology.* Ottawa, October 8-11, 2007. [Download PDF]

[2] Koller, S.M., Hardmeier, D., Michel, S., Schwaninger, A., (2008). Investigating training, transfer, and viewpoint effects resulting from recurrent CBT of X-ray image interpretation. *Journal of Transportation Security, 1*(2), 81-106. [Download PDF]

[3] Hardmeier, D., Hofer, F., & Schwaninger, A. (2006). The role of recurrent CBT for increasing aviation security screeners' visual knowledge and abilities needed in x-ray screening. *Proceedings of the 4th International Aviation Security Technology Symposium,* Washington, D.C., USA, November 27 – December 1, 2006, 338-342. [Download PDF]

[4] Schwaninger, A., Hardmeier, D., & Hofer, F. (2005). Aviation security screeners' visual abilities and visual knowledge measurement. *IEEE Aerospace and Electronic Systems, 20*(6), 29-35. [Download PDF]

[5] Schwaninger, A., Bridges, A., Drury, C., Durinckx, F., Durrant, P., Hodge, T., Hofer, F., Jongejan, R., Maguire, R., McClumpha, A., Neiderman, E., Steinmann, C., & Wüest, W. (2006). Principles and Requirements for Assessing X-ray Image Interpretation Competency of Aviation Security Screeners. *White Paper, International Aviation Security Human Factors Technical Advisory Group (InterTAG), Competency Assessment Working Group (CAWG).* [Request a copy]

[6] Hardmeier, D., Hofer, F., & Schwaninger, A. (2005). The X-ray object recognition test (X-Ray ORT) – a reliable and valid instrument for measuring visual abilities needed in X-ray screening. *IEEE ICCST Proceedings, 39*, 189-192. [Download PDF]

SECURITY IN PRACTICE USING THE CASRA TESTING PLATFORM (CTP)

Text: Christoph Meier & Sandrina Ritzmann

Tests are common in every area of life: They are necessary whenever we have to prove that we possess certain abilities, that we have learned something or that we are able to put newly acquired skills to practical use. In school, tests are used in order to check whether the learning content has been understood. Later, we have to take tests to get a driver's license, in the assessment process for certain jobs or even for a number of leisure activities. Based on scientific know-how and several projects with regulators, airports and security companies, the CASRA Testing Platform (CTP) was developed. It allows testing X-ray image interpretation competencies and the theoretical knowledge of screeners in an effective, efficient and flexible way.

First of all, it should be clarified why scientifically-based testing is so important to achieve and maintain the required standards in security screening. The answer is simple: scientifically-based tests can be used to decide whether an airport security officer (screener) is able and capable to do his job in an efficient and reliable way. When someone applies for a job as a screener, it is important to test, amongst other aspects, his visual abilities needed for X-ray image interpretation. Pre-employment assessment tests help selecting those people who have the potential to become good screeners. After the basic training and in regular intervals thereafter, tests of the acquired X-ray image interpretation competency and of theoretical and practical knowledge are important prerequisites to achieve and maintain the required security standards. For this reason, certification tests have been mandatory in Europe since 2010 and they have to be conducted at least every three years. In addition, internal quality control tests are useful in between certifications to monitor performance changes and react to them, for example by adjusting the organizational training requirements.

CASRA provides tests of high quality for these purposes. For pre-employment assessment, this is the X-Ray Object Recognition Test (X-Ray ORT). The X-Ray Competency Assessment Test (X-Ray CAT) has been designed for certification and internal quality control and is available for Cabin Baggage Screening (CBS), Hold Baggage Screening (HBS), cargo and mail screening. The X-Ray ORT and CAT are X-ray image interpretation tests. Furthermore, CASRA also provides multiple-choice knowledge tests, for example the Theoretical Exam on Computer (SecTEC). All CASRA tests are designed based on scientific knowledge and in close collaboration with regulators, airports and security companies. Extensive pre-study data is used to achieve high reliability, validity and standardization values[1]. The adherence to these so-called psychometric properties is important to assure that a test is "fair" in the sense that it does not produce mere chance results and that - based on the achieved scores - good screeners can indeed be distinguished from less competent screeners[2].

The second question that should be addressed is which form should be used to test X-ray image interpretation competencies as well as the theoretical knowledge. Quite often, tests are still conducted in paper and pencil format and the computer is used as data storage device only. However, several problems arise when a paper and pencil test is chosen. Showing X-ray images to a screener using paper and pencil, even if a projector is employed, is not very efficient. Furthermore, the test conditions are very different to the actual working conditions, where images have to be interpreted on a screen in a better quality and where image enhancement and zoom functions are available. Scientifically, a test not

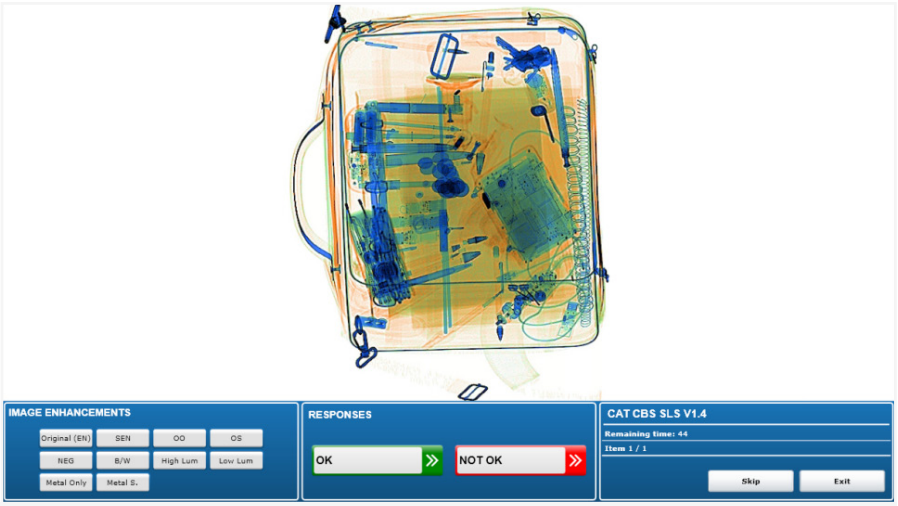


Figure 1: Example of an X-ray image interpretation test in the CTP

matching the operational conditions is said to have low "face validity". When it comes to theoretical knowledge tests, applying a multiple-choice test in a paper and pencil format is rather inefficient with respect to test correction and evaluation. Additionally, the correction of paper and pencil tests is error-prone.

In order to increase the efficiency and objectivity of tests, CASRA developed a computer-based testing platform which accounts for the needs of different stakeholders when it comes to testing. Ideally, a computer-based tool should cover four different scenarios regarding the implementation and/or development of X-ray image interpretation or theoretical tests like multiple-choice knowledge tests. The first scenario is the implementation of a standardized and validated test using pre-defined specifications. The second scenario is the implementation of a standardized and validated test, but with minor adjustments of parameters such as the viewing time of test images. The third scenario is that a new test is developed for an end-user according to specific requirements. In the fourth scenario, the end-user himself creates a tailor-made test covering specific situations. Numerous occasions are conceivable where the tailoring of tests is necessary. National regulations might, for example, specify certain test characteristics that need to be met by a certification test, such as the number of images in a test or the ratio of clear images to images containing a prohibited item. Other examples are the need for a test that contains only a certain type of threat, such as Improvised Explosive Devices (IEDs), or the need to adjust the performance criterion to pass a test to a certain population. Beside the different requirements of regulators or customers, CASRA's applied research projects in collaboration with regulators and industry partners often call for tests that are tailored to the questions and problems explored in the project.

COMPUTER-BASED TESTING OF SCREENERS

Based on these research and practical needs, the CASRA Testing Platform (CTP) was designed and developed as a generic tool for X-ray image interpretation tests and multiple-choice knowledge tests (see Figure 1 & 2) in order to guarantee flexibility and customization.

To cover the first scenario described above, the CTP allows running the X-Ray ORT, X-Ray CAT and the SecTEC. As required in the other scenarios, several parameters that emerged to be relevant for regulators, airports and security service providers or in research projects can be configured to meet specific needs and requirements as well as to guarantee the face validity of a test. Some examples include the following:

- The **performance criterion** to pass a test (the so-called passmark) can be configured, for example if the same test is used for novices and expert

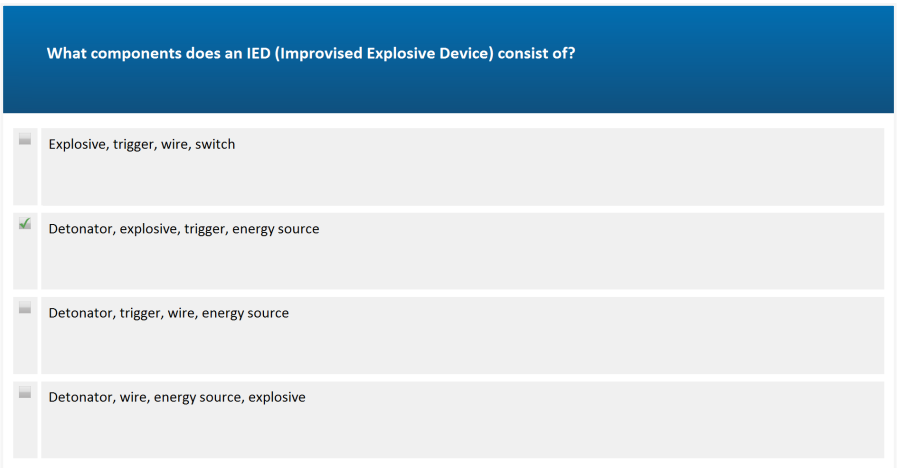


Figure 2: Example of an item in the SecTEC

screeners, but with different performance level requirements. The passmark value is taken into account for test feedbacks and test reports.

- The **viewing time** for the images in a test can be changed. This option can be useful if a test has to be adapted to the training level of test-takers (e.g. novices are granted more viewing time) or to the X-ray screening reality of the examinees (e.g. cabin-baggage screening at an airport has to be performed much quicker than X-ray screening of bags in a prison).
- Another configurable feature is the possibility to combine different tests to a **test series** (e.g. if some test-takers have to take a CBS and an HBS image interpretation test, optionally in combination with a multiple-choice knowledge test). Test series can be allocated to certain persons or groups of personnel. When an examinee starts the test series, all tests in the series are shown consecutively. The access to test series can be limited by a password feature or by activating the test only for a certain period of time. This can be useful to establish a more sophisticated quality control scheme where performance data is collected in a pre-defined time frame.
- Depending on the purpose of a test, a different number of **allowed attempts** per examinee can be activated. Up to the configured number of attempts, the examinees can independently access the test again.

Furthermore, the CTP allows generating **new and individualized tests** so that CASRA or airports, security service providers and national regulators can create and configure tailored X-ray image interpretation tests as described in the third and fourth scenario.

- Administrators are free to **choose any number and type of images** from an image library. The ratio of images that are clear and images that contain a threat can be individually adjusted.
- By default, the **answer categories to judge an image** are set to OK and NOT OK, but they can be adapted if required in a test (for example to CLEAR BAG and SEARCH).
- Furthermore, it can be decided whether an **individualized test completion feedback** is given upon finishing a test and what information it shall contain.
- With the **image uploader functionality**, administrators can upload their own images and rename them as they wish to integrate them in their tests. Each image has to be defined as containing a threat or not. For threat images, the category of the threat (for example "knife" or "IED") can be chosen and the location of the potential threat can be marked.

To sum up, the CTP is a tool that covers all four testing scenarios, from the use of a standardized test to the creation of a fully customized test. But as with any tool, it is in the user's responsibility to prudently apply it. Individualized tests have a number of advantages, but they also come with certain disadvantages.

ADVANTAGES AND DISADVANTAGES OF INDIVIDUALLY CREATED TESTS

On the one hand, individually created tests offer the possibility to quickly include **new threats** in a test or adapt to new regulations. Furthermore, **specific images** can be recorded to cover aspects such as the location of a threat (e.g. the placement of drugs in customs X-ray screening) or a particular type of threat (e.g. only a certain type of blasting composition in IEDs). Moreover, the **length of a test** can be determined to match its purpose: An internal quality control test that should give an organization a rough idea of the competency level of the screener workforce could be shorter than a certification test that is linked to the official authorization to work as a screener. Additionally, the **test difficulty** can be varied and changed if necessary. Hence, tests can be adjusted to the operational reality and requirements, which is a prerequisite for the validity of a test.

On the other hand, the preparation of an individual test is time-consuming and requires considerable knowledge

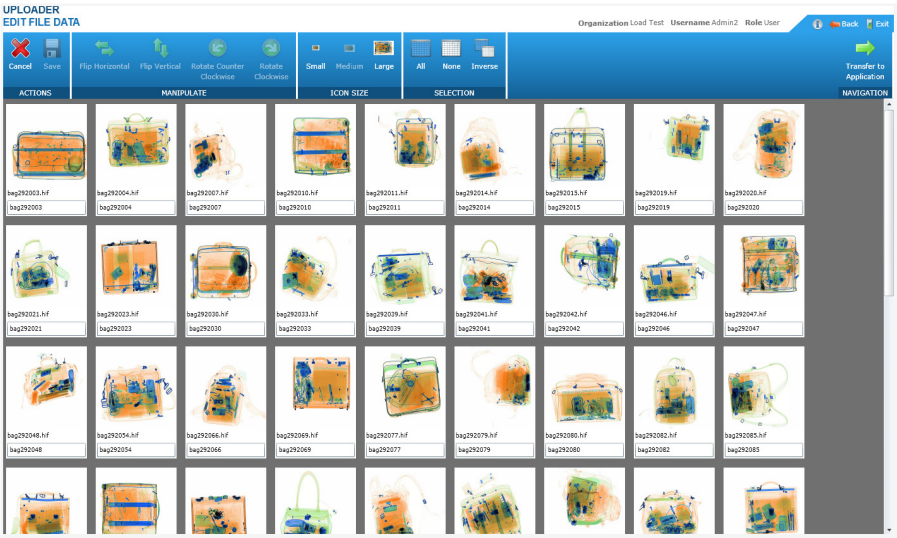


Figure 3: Screenshot of the Image Uploader module in the CTP

of scientific test design. If tests are created without taking into account psychometric principles, **sufficient reliability, validity and standardization cannot be guaranteed**. In a science-based X-ray image interpretation test, the difficulty and content of test images should be carefully balanced to assure that the test is fair and well-founded. In the fourth scenario, where end-users create their own tests, they should make sure that psychometric know-how is considered to assure the quality of a test. Another aspect that has to be given sufficient consideration is the **choice of the passmark**. Ideally, a passmark is defined in a **pilot study** to find the value that, in a given population, best distinguishes between competent and less competent screeners. The introduction of certification tests should thus be accompanied by a research project to evaluate different passmark scenarios[2]. The goal should be to find a challenging, yet realistic passmark.

To sum up, the CTP was created to support the important and partially even mandatory testing of competencies in the area of X-ray image interpretation and theoretical knowledge. It is a computer-based tool that supports the efficient and objective use and creation of tests and that covers the four described scenarios of X-ray image interpretation testing, from the most basic one where an existing test is implemented to the most sophisticated one where an end-user uploads images and creates his own test. Therefore, the CTP is a tool that offers high flexibility and that can be applied for many purposes in different situations.

Further information on the CTP can be requested from info@casra.ch.

[1] This procedure is in accordance with the ECAC White Paper on the assessment of X-ray image interpretation competency: Schwaninger, A., Bridges, A., Drury, C., Durinckx, F., Durrant, P., Hodge, T., Hofer, F., Jongejan, R., Maguire, R., McClumpha, A., Neiderman, E., Steinmann, C., & Wüest, W. (2006). Principles and Requirements for Assessing X-ray Image Interpretation Competency of Aviation Security Screeners. White Paper, International Aviation Security Human Factors Technical Advisory Group (InterTAG), Competency Assessment Working Group (CAWG).

[2] More information on scientific test design and the importance of a careful definition of the passmark can be found in our Newsletter No. 1, which can be requested from info@casra.ch.

Science & Security – CASRA Newsletter
ISSN 1664-5715
Published three times a year

CASRA
Thurgauerstrasse 39
8050 Zurich, Switzerland

Phone +41 (0)43 336 01 01
Fax +41 (0)43 336 01 00
E-mail info@casra.ch
Web www.casra.ch