SCIENCE & SECURITY July 2012

CASRA NEWSLETTER – ISSUE 1

The Center for Adaptive Security Research and Applications (CASRA) was founded in 2008 and aims at increasing security and facilitation at airports and in other environments involving people and technology. CASRA uses an adaptive approach by combining applied psychology, computer science and economic analyses.

With this newsletter, which will be distributed three times a year, we want to make scientific evidence available to the public and show how it can be put into practice.

We will single out topics that seem enriching and inspiring and publish them in two sections: "Research put across" and "Security in practice".

The aim of the newsletter is furthermore to create an active platform for the exchange of experiences between research and practice.

We hope you enjoy reading the news and we will be happy to receive your feedback at info@casra.ch.

A. Schwaninger

Prof. Dr. Adrian Schwaninger

Dr. Diana Hardmeier

TOPICS IN THIS ISSUE:

RESEARCH PUT ACROSS

FROM VISION RESEARCH TO A WIDELY USED TOOL FOR SELECTION: IMAGE-BASED FACTORS AND THE X-RAY OBJECT RECOGNITION TEST

Not every person can become a good and efficient security screener. The questions relevant for personnel selection are: which abilities are needed for efficient visual inspection of X-ray images, and how can they be tested before employing someone? Vision research has led to important insights which resulted in a widely used test for pre-employment assessment and selection.

SECURITY IN PRACTICE

SECURITY OFFICER CERTIFICATION: "MORE TO IT THAN MEETS THE EYE"

Security officer certification is legally required in the EU, confronting member states with the challenge to design certification tests and implement a certification process. Three steps for test design and a range of best practices ensure the adherence to high scientific and psychological standards.

FROM VISION RESEARCH TO A WIDELY USED TOOL FOR SELECTION: IMAGE-BASED FACTORS AND THE X-RAY OBJECT RECOGNITION TEST

Text: Alex Kunz, Sandrina Ritzmann & Adrian Schwaninger

It is a well-known fact that not every person can become a good pilot. However, many people do not know that it is similar for security officers working with X-ray machines (X-ray screeners). Research has shown that not everyone can become a good X-ray screener. The questions relevant for personnel selection are: which abilities are needed for efficient visual inspection of X-ray images and how can they be tested before employing someone?

To answer these questions, researchers from CASRA have carried out several research studies since 2003. Based on scientific findings on object recognition and visual cognition, **three so-called image-based factors were discovered that influence the detection of a prohibited item in a bag** ([1], [2], [3]). Figure 1 illustrates these image-based factors. The first factor is viewpoint. When objects are rotated and depicted from an unusual viewpoint, they can become difficult to recognize. Another factor is superposition. Objects in X-ray images are often superimposed by other objects, which also has an effect on detection performance: the more superposition, the harder it is to identify an object. The third factor is bag complexity, which depends on the number and type of objects in a bag. Prohibited items are more difficult to detect when they are in a densely packed bag, as other objects distract attention. These image-based factors represent challenges that are inherent to the task of security screeners – unusual viewpoints of objects, superposition, and bag complexity cannot be avoided in day-to-day operations.

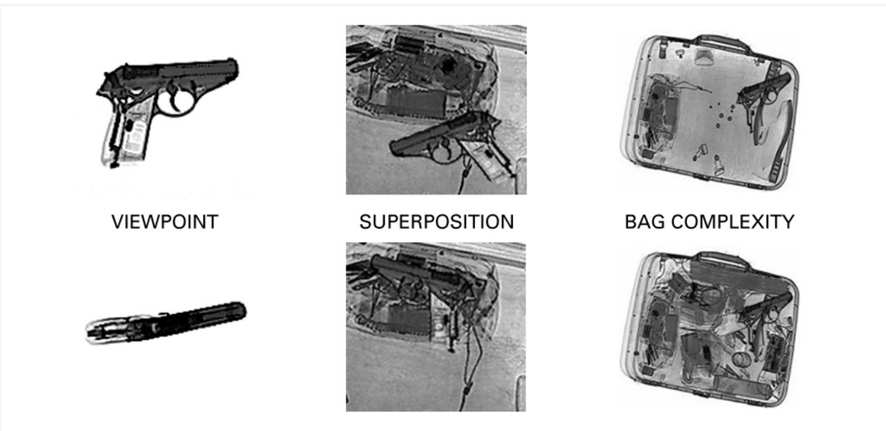

Figure 1: Image-based factors

The next step after the discovery of the image-based factors was to explore whether people differed in their abilities to cope with these factors. To this end, the X-Ray Object Recognition Test (X-Ray ORT) was developed ([1], [3]). The X-Ray ORT is a computer-based test and consists of 256 X-ray images. Only guns and knives have to be identified, due to the fact that the shape and appearance of these objects are known by most people, even if they have never worked with X-ray images before. Furthermore, all images are shown in black and white, since job applicants do not know the meaning of the different colors of real X-ray images. To measure the visual abilities of test candidates systematically, rotations of guns and knives showing familiar and unfamiliar viewpoints are used with little and much superposition by other objects in bags of low and high levels of bag complexity. In the test, each image is shown for four seconds on the computer screen because at rush hour, X-ray screeners often have only three to five seconds to visually inspect X-ray images. For each X-ray image, the test candidate has to decide whether the bag is OK (contains no gun and no knife) or NOT OK (contains a gun or a knife)[1].

For the X-Ray ORT to be a valuable tool for pre-employment selection, it would have to reveal differences in the individual abilities of job applicants to cope with the image-based factors so that the best candidates can be chosen. **A study with 284 screener candidates indeed showed prominent differences between people regarding their abilities to cope with effects of viewpoint, superposition and bag complexity.** The variance of individual test results was large, represented by the thin vertical lines in figure 2, which depict the standard deviation (a measure of variance between individuals), and the stars, which show the 2.5% best and the 2.5% worst performing individuals. The effects of viewpoint, superposition, and bag complexity were analyzed to investigate whether the X-ray images in the test accurately depict them. Results indeed showed systematic effects of the image-based factors. Showing threat items in rotated view led to a significantly lower detection performance. When guns and knives were superimposed by other objects, recognition performance dropped substantially. Finally, prohibited items were harder to detect in more complex compared to less complex bags (see figure 2). Similar results indicating differences in the abilities of test takers and effects of the image-based factors were shown in other studies with novices and experienced screeners from several European airports ([4]). Even for screeners with several years of working experience, large differences could be found with regard to how well they coped with the effects of viewpoint, superposition, and bag complexity ([2]).

To sum up, the X-Ray ORT is an instrument that can be used to assess the abilities of job applicants to cope with image-based factors in X-ray images. It allows identifying the most capable candidates who show high potential to become competent screeners. This notion is further supported by data showing that the results of the X-Ray ORT were significantly related to the detection performance measured with Threat Image Projection (TIP) on the job – the higher the screeners' score in the X-Ray ORT, the more threat items they discovered at work ([4]).

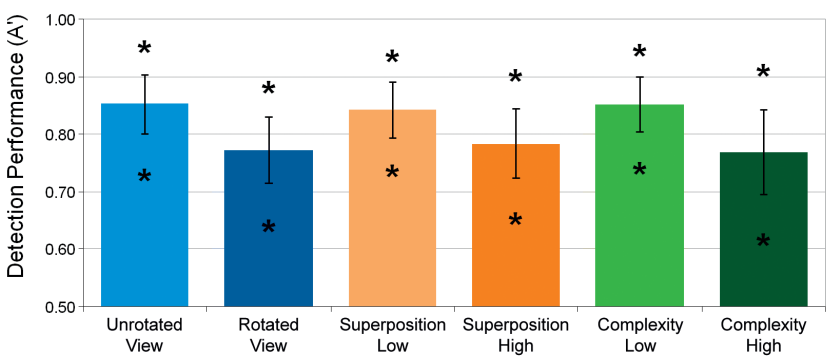

Figure 2: Effects of viewpoint, superposition, and bag complexity (study with 284 applicants for a security screening job) on detection performance (A')[2]. Thin vertical lines represent standard deviations. Stars indicate the scores of the 2.5% best and 2.5% worst performing candidates.

At the beginning of this article it was asked which abilities are needed to inspect X-ray images and how they could be tested in job candidates. The presented results showed that X-ray screeners must be able to cope with viewpoint, superposition, and bag complexity, and that the X-Ray ORT is a scientifically reliable, valid and standardized instrument to select the job candidates with the highest potential to meet this requirement. For further information on pre-employment assessment tools or a free X-Ray ORT trial (one test per person), please contact us.

REFERENCES

[1] Schwaninger, A. (2003). Evaluation and selection of airport security screeners. *AIRPORT, 2003*(2), 14-15.

[2] 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.

[3] 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.

[4] Schwaninger, A. (2005). Increasing efficiency in airport security screening. *WIT Transactions on the Built Environment, 82,* 407-416.

[5] Hofer, F., & Schwaninger, A. (2004). Reliable and valid measures of threat detection performance in X-ray screening, *IEEE ICCST Proceedings, 38,* 303-308.

[6] J.B. Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75,* 424-429.

[7] Pastore, R. E., Crawley, E. J., Berens, M.S., & Skelly, M. A. (2003). "Nonparametric" A' and other modern misconceptions about signal detection theory. *Psychonomic Bulletin & Review, 10*(3), 556-569.

[1] For a more detailed description of the test design see reference [3].

[2] A' is a measure of detection performance. It takes hit rate and false alarm rate into account (for details regarding A' see references [5], [6], and [7]).

SECURITY OFFICER CERTIFICATION: "MORE TO IT THAN MEETS THE EYE"

Text: Sandrina Ritzmann & Adrian Schwaninger

Today's complex air transport system is heavily dependent on modern and advanced technology. This is true for all aspects of the system, including airport security. Developments in X-ray screening equipment for passenger baggage, but also for cargo and mail in the last decade have resulted in images of increasingly high quality. But even the best technology is ineffective without competent personnel who can put its functions to use. The human factors, i.e. the skills and abilities of security officers to interpret X-ray images, are thus of high importance for the success of the whole aviation security system. A means to ensure that security officers indeed possess all the necessary skills to fulfill their tasks are regular competency assessments in the form of initial and recurrent certification tests.

Awareness of the importance of competency assessments and certification testing has grown considerably in the last few years. Consequently, certification is now a legal requirement in the European Union: EU regulation (EC) 300/2008 and the commission regulations (EU) 185/2010 demand mandatory initial certification of security officers and recurrent certification at least every three years. Since the coming into effect of these regulations in 2010, a transition period for the introduction of national certification programs has been granted, but compliance will be monitored as from 2013. Member states of the European Union and states that adopted European Regulations have thus been faced with the challenge of designing and setting up certification tests and processes. **The goal of this article is to give a short overview on the necessary steps in the design of certification tests from a scientific and psychological point of view.** Moreover, **best practices** for security officer certification are described.

DESIGN OF CERTIFICATION TESTS: THREE STEPS

The design of a test of X-ray image interpretation competency should ideally be carried out in three steps. **The first step should be to determine what task the examinee has to be able to accomplish.** The main task of security officers is to detect prohibited items in X-ray images of passenger baggage, cargo, or mail. Therefore, a certification test should measure the detection performance of the test taker. The measures that define the detection performance are hit rates and false alarm rates. When a screener correctly reports that an X-ray image contains a prohibited item (i.e. a threat), the response is counted as a hit. However, if a clear image not comprising any prohibited item is reported to contain a threat, the response is categorized as false alarm. Although the main task in X-ray screening is the detection of threat items, an additional option could be to include certain non-threat objects in the test and require the candidates to identify and/or name them.

The second step in the design of a certification test is to make sure that the test is "fair" and does not produce mere chance results. From a psychological point of view, a fair test has to fulfill a range of criteria, also called psychometric properties: It should be reliable, valid, and standardized. To ensure that these three criteria are met, theoretical considerations and psychological know-how regarding test design are of high importance. It is for example advisable to use the statistical measure A' that is calculated based on hit and false alarm rates to quantify detection performance. One advantage of using A' as a single performance measure is that it facilitates standardized interpretation of the test result. Please refer to the infobox in this article for a more detailed description of reliability, validity, and standardization.

The third and final step of test design is to define the detection performance that security officers need to achieve to pass the certification test. To define a minimum performance standard, it seems logical to start with theoretically founded requirements and set the stipulated performance level accordingly (similar to, e.g., language certificates). However, with regard to security officer certification, this approach bears the risk that security officers, not used to being tested and certified, might not be able to show their full potential from the beginning. Consequently, they might initially fall short of the requirements due to the novelty of the process or a lack of specific training. To avoid this, data from a pilot study and the norm established therein (see section on standardization in the infobox) should be used to determine the competencies and the actual performance level of the security officer population before the certification process is made compulsory. Based on these results, challenging yet realistic performance level requirements can be set.

BEST PRACTICES FOR SECURITY OFFICER CERTIFICATION

For the process of security officer certification, not only the three steps described above, but also a range of best practices are important. They are described in more detail below:

Use Computer-based Tests

The most recommendable method to assess X-ray image interpretation competency are computer-based tests. Although covert testing by inspectors and threat image projection (TIP) can be used as well, these methods are much more challenging to design so that they are reliable, valid, and standardized. Computer-based tests should consist of X-ray images of bags, cargo, or mail with and without prohibited objects that are displayed for a duration comparable to operational conditions. For each image, the test candidate should indicate whether or not a threat is present. The ratio of clear images and images containing a threat can be varied. The resulting test data should be maximally informative and satisfy all the requirements for sound statistical analyses.

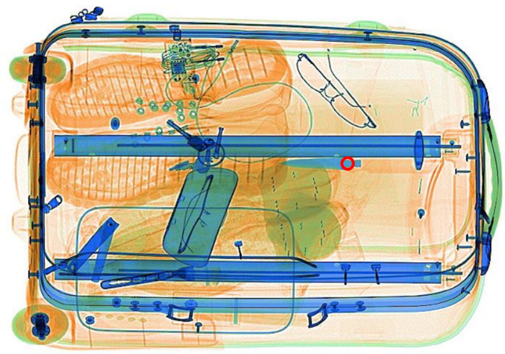

Figure 1: Example test image from the X-Ray Competency Assessment Test (CAT): a reliable, valid, and standardized certification test developed by CASRA.

Carefully Design the Test Images

Research has shown that the probability of detection of a prohibited item depends not only on the knowledge and competencies of the screener, but also on the general difficulty of the presented threat item. Furthermore, image-based factors influence detection performance (see also article on pre-employment selection in this issue): These are rotation of the prohibited item (view difficulty), superposition by other objects, and number and type of other presented objects (bag complexity). It is thus necessary to carefully design the test images, taking into account and balancing their general difficulty as well as the image-based factors (see figure 1 for an example). Furthermore, it is important to update the test images regularly to ensure that the test content remains up-to-date with current threats, X-ray image quality standards, and regulatory requirements. For example in Switzerland, new test images are constructed and validated for every cycle of recurrent certification, which is due every two years according to the National Aviation Security Program.

Harmonize the National Certification Process

While the design of the test and test images involves many decisions on a micro level, macro level decisions influence the success of the certification process as well. Six years of experience with certification in Switzerland, where CASRA develops and conducts certification tests on behalf of the appropriate authority, have shown that a nationally harmonized certification process has clear advantages. Harmonization includes that nationally, a single, standardized test is used. This leads to results that are comparable among different airports and security service providers. Furthermore, the process should be coordinated on a national level, reducing administrative effort. Finally, the data analysis should be scientifically founded and centralized to ensure the adherence to high statistical and psychological standards.

Train the Competencies That You Test

Although training and certification of security officers should contain different x-ray images and test questions (otherwise, the criticism of "teaching to the test" could be raised), they should both be based on and oriented towards the same set of competencies necessary for a security officer to show high detection performance. This does not only make sense from a psychological point of view, but also from a practical perspective, as it would not be beneficial for security if screeners were trained in skills they do not need on-the-job. Furthermore, regular training is vital for the maintenance of skills between tests, especially when it comes to the detection of rare threat items such as improvised explosive devices (IEDs).

In summary, this short overview on test design and best practices regarding security officer certification has shown the **importance of psychological know-how when certification tests are developed and a certification process is introduced.** Although only a few of the important aspects could be highlighted, it can be concluded that "there is more to certification than meets the eye". If you have any further questions regarding the content of this article or wish to obtain more detailed information, please do not hesitate to contact us.

RELIABILITY

The reliability of a test can be described as its "consistency" or "repeatability". Consistency means that the items of a test should all measure the same competency – in the case of security officer certification, X-ray image competency. Repeatability means that the test should yield approximately the same result today and on a later test occasion (given that the security officers did not do any training between tests). Statistically, there are several methods to determine a test's reliability. The split-half reliability is calculated as the correlation (i.e., the relationship) between one half of the test and the other half. The stronger the relationship is, the more reliable is the test, meaning that examinees get approximately the same score in both halves. Another method is the test-retest-reliability. Here, the relationship between the test and a repetition of the test after a certain time interval is calculated. A third method is to calculate the internal consistency of a test (for example using a measure called "Cronbach's Alpha"), which takes the relationships of all test items with each other into account. **A test is usually judged as acceptable when reliability coefficients lie above 0.7. To reliably measure individual performance, correlation coefficients of at least 0.75 and Cronbach's Alpha of at least 0.85 are recommended.**

VALIDITY

Validity describes the property of a test that it measures what it is supposed to measure. Security officer certification tests should measure detection performance. This can only be done in a valid way if both hit rate and false alarm rate are taken into account. The reason is that a high hit rate could be achieved either by correctly singling out only the X-ray images containing prohibited items, as a competent screener would do, or by simply judging most images as "not OK". The latter strategy is not desirable as it would lead to a high false alarm rate. However, it could not be identified with a test measuring only the hit rate – the two strategies would not be distinguishable. Such a test would not be valid, as it would not differentiate between more competent and less competent, or trained and untrained security officers.

STANDARDIZATION

The third and last criterion for a test is standardization. **Standardization is realized when the testing procedure and the test analysis follow given rules and are objective and independent of the test environment and the test instructor.** Examples are the consistent interpretation of scores (e.g., it is clearly defined when a score can be judged as "above average") or standardized test instructions (every instructor uses the same wording to instruct the examinees). Standardization also involves the establishment of a "norm" of the relevant comparison group (e.g., the security officers of the respective country). This includes the determination of the group's average performance in a pilot study, so that individual scores can be compared to it and judged as below or above average. To facilitate the standardized interpretation of test results and make them more comparable, it is highly advisable to combine hit and false alarm rates into one single statistical measure, for example A' (see references [5], [6] and [7] in the article on selection in this newsletter for more information on A').

FURTHER READING

Principles and requirements for assessing X-ray image interpretation competency (White Paper):

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]

Description of a reliable, valid and standardized certification test (X-Ray CAT):

Koller, S., & Schwaninger, A. (2006). Assessing X-ray image interpretation competency of airport security screeners. Proceedings of the 2nd International Conference on Research in Air Transportation, ICRAT 2006, Belgrade, Serbia and Montenegro, June 24-28, 2006 , 399-402. [Download PDF]

IMPRESSUM

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

Editors:
Prof. Dr. Adrian Schwaninger
Dr. Diana Hardmeier

CASRA/APSS
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