SCIENCE & SECURITY Issue 25

CASRA NEWSLETTER

It is time to publish our first newsletter in 2021!

In this 25th issue, the section “Research Put Across” introduces CASRA’s investigation of human-machine interaction between airport security screeners and automated explosives detection systems for cabin baggage (EDSCB). In the “Security in Practice” section, we present the results of the International Cargo Screening Study, which was conducted together with the International Air Transport Assosiation (IATA).

As always, we are looking forward to receiving any feedback you might have as well as your input on topics you would like us to address in upcoming newsletters.

We hope you enjoy the insight given through our newest articles.

With best wishes,

Dr. Diana Hardmeier
Director

Prof. Dr. Adrian Schwaninger
Chairman

TOPICS IN THIS ISSUE:

RESEARCH PUT ACROSS

AUTOMATION RELIABILITY, HUMAN–MACHINE SYSTEM PERFORMANCE, AND OPERATOR COMPLIANCE

The following article covers the CASRA’s investigation of human-machine interaction investigation between airport security screeners and automated explosives detection systems for cabin baggage (EDSCB). The scope of the study was to understand how automation reliability and unaided performance relate to human-machine system performance and operator compliance.

SECURITY IN PRACTICE

INTERNATIONAL CARGO SCREENING STUDY

In collaboration with the International Air Transport Association (IATA), and supported by id3 Technologies as well as participants from all over the world, CASRA is delighted to introduce a web-based international cargo screening study. The aim of the study is to illustrate the benefits of computer-based training (CBT) for cargo screening and show its potential for improvements.

CASRA

AUTOMATION RELIABILITY, HUMAN–MACHINE SYSTEM PERFORMANCE, AND OPERATOR COMPLIANCE: A STUDY WITH AIRPORT SECURITY SCREENERS SUPPORTED BY AUTOMATED EXPLOSIVES DETECTION SYSTEMS FOR CABIN BAGGAGE SCREENING

Text: David Huegli

Automation represents functions performed by machines that assist or replace tasks usually executed by humans [1]. The benefits of automation depend on the automation reliability or the unaided performance of the human operators [2]. System designers must decide how to allocate functions of automation to achieve an optimal human-machine match. At CASRA, we investigated the human-machine interaction between airport security screeners and automated explosives detection systems for cabin baggage (EDSCB). Our study allowed us to understand how automation reliability and unaided performance relate to human-machine system performance and operator compliance. We derive recommendations for the use of EDSCB in X-ray screening of cabin baggage at airports.

KEY MESSAGES: AUTOMATED EXPLOSIVES DETECTION SYSTEMS FOR CABIN BAGGAGE (EDSCB)

- The benefits of automation depend on automation reliability and unaided performance
- We illustrated issues coming up with function allocation between humans and automation in airport security X-ray screening.
- Three different EDSCB with on-screen alarm resolution were compared
- When unaided performance is low, EDSCB with a high reliability increases target detection
- Operator compliance depends on the probability of a target being present when EDSCB alarms
- Operators should be given clear instructions if the probability that alarmed bags contain explosives is below 50%.

AUTOMATION RELIABILITY AND HUMAN-MACHINE SYSTEM PERFORMANCE

Various taxonomies distinguish different levels of automation [3]–[5]. One form of low-level automation provides alarms to support human operators by cueing attention to areas of a display that might contain a target. Examples include air-traffic control conflict alerting systems [6], combat systems for better detection of enemy soldiers on the battlefield [7], support systems for medical diagnoses [8], and automated EDSCB to assist X-ray screeners to search passenger baggage [9]–[11].

The benefits of low-level automation depend on unaided performance and automation reliability (see [2], for a review). When operators can perform well without automation, their self-confidence increases [12], and they rely more on their own judgment than on what automation suggests [13], [14]. When unaided performance is lower, operators' self-confidence is also lower [12], and operators benefit more from automation if reliable [15], [16].

Low-level automation systems perform tasks for which unreliability and imperfection are expected (even if it is not desirable) because of technical reasons. A small number of imperfections such as automation misses (a target is not detected by automation) or false alerts (alarm in the absence of target) are tolerated [2]. System designers must decide how to set the automation's detection thresholds. Usually, a higher priority is given to reducing automation misses at the expense of accepting more automation false alarms [17]. However, if operators experience many false alarms, this can result in a *cry-wolf effect,* with operators ignoring automation alarms [18]. Several authors have described the *cry-wolf effect* as a lack of operator compliance (e.g., [15], [19], [20]). Compliance is the degree to which operators act in accordance with the automation alarm and represents the active processing of automation alarms. It is well established that automation reliability affects operator compliance. Specifically, compliance with automation depends on the positive predictive value (PPV) of automation. The PPV is the probability of a target being present when automation alarms [21]–[23]. Several studies have shown that compliance with automation increases when PPV increases [16], [22], [24]–[28].

AIRPORT SECURITY X-RAY SCREENING OF CABIN BAGGAGE

Automation can be used to support screeners during the visual inspection of X-ray images at airport security checkpoints. Several studies have consistently shown that computer-based training is essential to achieve and maintain good detection performance of screeners [29]–[32]. Bombs, the most dangerous prohibited articles in passenger baggage, are technically called improvised explosive devices (IEDs). They are composed of a power source, a triggering device, a detonator, and an explosive charge connected by wires [33], [34]. Screeners can learn to recognize these components with computer-based training and they can achieve and maintain a high detection performance for IEDs [29]–[32], [35]. In cabin baggage screening, bare explosives also pose a threat because these could be combined with other IED components after passing an airport security checkpoint [10]. The detection of bare explosives can be a challenge even for well-trained screeners because they lack the other components of IEDs (power source, triggering device, and detonator) and often look like a harmless organic mass [36], [37]. Automation EDSCB indicate areas in an X-ray image that might contain explosives in order to support the operator to detect such explosives, i.e. IEDs and bare explosives, in cabin baggage screening. EDSCB can help to overcome the difficulties in detecting explosive material.

At airport security checkpoints, the prevalence of prohibited articles in X-ray screening (especially explosive material) is scarce. However, the incidence of prohibited articles (guns, explosives, knives, and other threat items) is increased artificially to about 2–4% using threat image projection (TIP), a technology that projects prerecorded X-ray images of prohibited articles into X-ray images of real passenger bags [38]–[41]. Furthermore, airports, government authorities, and security companies conduct covert tests at security checkpoints using prohibited articles, including guns, simulated explosives, knives, and other threat items [42], [43]. With TIP and covert tests, the target occurence of (simulated) explosives in X-ray images is about 1–2%. Thus, when a screener works at an airport security checkpoint and she or he visually inspects a hundred X-ray images, only one to two will contain (simulated) explosives.

EXLOSIVES DETECTION SYSTEMS FOR CABIN BAGGAGE (EDSCB)

European regulation standards require EDSCB machines to achieve particular reliability [44]. The hit rate of the EDSCB should not only be high enough to deter terrorists, but also to reduce misses that lead to disastrous consequences. However, the increased hit rate comes at the cost of automation false alarm rates of 6–20% [10], [45]. As a result, the EDSCB yield 6–20 alarms of each 100 bags – and most of them are false alarms.

In most countries, airports use EDSCB with an automated decision (high-level automation). In this case bags that are alarmed by the EDSCB to be suspicious of possibly containing explosive material are automatically sent to secondary search, where a second screener searches the bag manually.

Regarding checkpoint efficiency, the human-machine system false alarm rates of baggage screening should be minimal, which means, 20% of all bags sent to secondary search is impractical. Further examination of bags at the checkpoint involves more time-consuming explosives trace detection and the manual opening of baggage which lowers checkpoint throughput [45].

However, some countries use EDSCB with on-screen alarm resolution (low-level automation) to reduce the false alarm rate of the system. EDSCB provide screeners with red frames around potential threats on the X-ray image. They must decide whether the baggage contains an actual threat or not, and if a threat is detected, send it to secondary search. The goal for airports should entail preventing screeners from dismissing true EDSCB alarms and keep the reject rates in a practically feasible range. Unfortunately, EDSCB false alarms pose a problem for human-machine system performance and operator compliance resulting in a *cry-wolf effect*. In this study, we investigated EDSCB implemented as low-level automation.

OUR STUDY

At CASRA, we conducted a study with a simulated cabin baggage screening task in which screeners had to detect guns, knives, IEDs, and bare explosives on dual-view X-ray images. EDSCB supported screeners by indicating areas in X-ray images that might contain IEDs or bare explosives. See Figure 1 for the interface of our simulated cabin baggage screening task. We simulated three different EDSCB scenarios that were inspired by discussions with appropriate authorities responsible for national aviation security regulations on X-ray screening equipment and security control procedures at airports. The authorities were interested in knowing whether the large *cry-wolf effect* for detecting bare explosives found by a previous study with single-view EDSCB [10] also applies to the multi-view EDSCB more commonly used at airports. Further, we wanted to investigate whether using EDSCB with a lower false alarm rate would improve human-machine system performance.

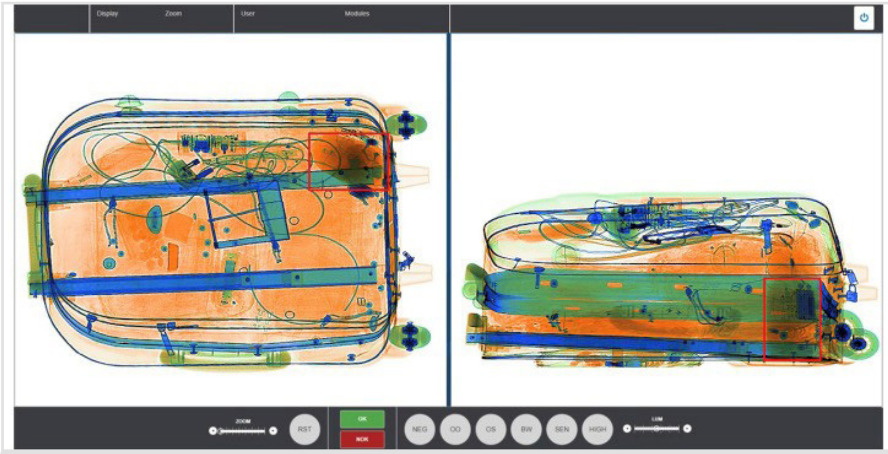


Figure 1: The interface of the X-ray machine simulator used in our study. Red frames indicate areas that might contain explosive material (automation alarms). The same bag is shown from two viewpoints differing by about 90°.

STATE-OF-THE-ART EDSCB

Our *State-of-the-art EDSCB* had similar reliability as the ones currently employed at checkpoints with multi-view X-ray machines (hit rate 86%, false alarm rate 20%). As mentioned above, issues now discussed with EDSCB mainly concern the false alarm rate of these systems. EDSCB machines that achieve a C2-Standard according to the European regulations [44] have a relatively high automation hit rate of 80-90%. However, this comes with a relatively high false alarm rate of about 20%. When EDSCB is used as high-level automation and bags are automatically sent to secondary search, the false alarm rate of the system exceeds a practically feasible threshold. When EDSCB is used as low-level automation with on-screen alarm resolution, the screener could reduce the false alarm rate substantially. *State-of-the-art EDSCB* was considered as EDSCB with high automation reliability and it had a PPV of .3.

LOW FALSE ALARM RATE EDSCB

Another possibility would be an EDSCB with a hit rate of 60%, which reduces the false alarm rate to about 5%. Such an EDSCB might be enough to deter terrorists, considering that they have often used attack methods with a low risk of being caught [46]. Moreover, computer simulations [45] indicated that EDSCB with low false alarm rates of around 5% would allow resolving all automation alarms with explosive trace detection without negative impacts on screening efficiency (passenger and baggage throughput). Our *Low false alarm rate EDSCB* had the same automation reliability as *State-of-the-art EDSCB* but a lower automation hit rate and a low automation false alarm rate (hit rate 60%, false alarm rate 5%). *Low false alarm rate EDSCB* was also considered as EDSCB with high automation reliability. However, whereas *State-of-the-art EDSCB* had a PPV of .3, *low false alarm rate EDSCB* had a PPV of .5.

RISK BASED EDSCB

A third possibility would be a risk based screening scenario where *State-of-the-art EDSCB* is activated for only one-quarter of the screened bags based on passenger and flight-risk information. The unpredictable variation maintains deterrence. Our third EDSCB simulated a risk based screening scenario (hit rate 21%, false alarm rate 5%). It was considered an EDSCB with low automation reliability. The PPV of *Risk based EDSCB* was .3 and had, therefore, the same PPV as *State-of-the-art EDSCB.*

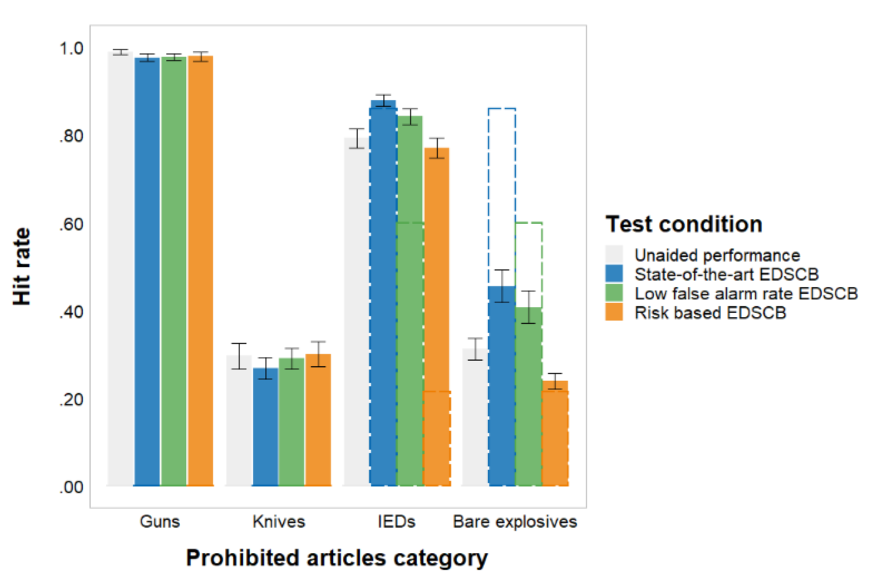


Figure 2: Human–machine system hit rate by test condition and prohibited articles category (IEDs, bare explosives, guns, and knives). Error bars show ±1 standard error of the mean. Dashed lines indicate the hit rates of automation by itself (automation hit rate).

EFFECTIVENESS DEPENDS ON AUTOMATION RELIABILITY…

Figure 2 shows that overall guns and IEDs were detected well in our study, whereas bare explosives and knives were more challenging to detect. Similar results have been found in previous studies for well-trained screeners without automation support [29], [30], [35] and in a recent study that tested screeners supported by single-view EDSCB [10]. Consistent with previous research on low-level automation conducted with student participants (see [2], for a review), we found that the benefits of automation depend on unaided performance and automation reliability. Overall, the human-machine system hit rate was higher when using automation with high reliability (*State-of-the-art EDSCB* and *Low false alarm rate EDSCB*) than an automation with low reliability (*Risk based EDSCB*), providing strong evidence for the relevance of high automation reliability.

… AND UNAIDED PERFORMANCE

Furthermore, our results confirm that when operators perform well without automation (in this study IEDs) it increases their self-confidence [12] and they rely more on their own judgment than on what automation suggests [13], [14]. We found high operator confidence (see Figure 3) and a high human-machine system hit rate between 79% and 88% for detecting IEDs, which was much higher than automation by itself in the two conditions with lower automation hit rates (21% in the *Risk based EDSCB,* and 61% in the *Low false alarm rate EDSCB*). For detecting difficult targets (bare explosives), unaided performance and screener confidence were lower (see

Figure 3) and automation provided large benefits in the two conditions with higher automation reliability (*State-of-the-art EDSCB* and *Low false alarm rate EDSCB*).

THE CRY-WOLF EFFECT

However, especially when working with the *State-of-the-art EDSCB*, screeners ignored many true EDSCB alarms. Therefore, human-machine system performance was lower than automation by itself (86% in the *State-of-the-art EDSCB* and 60% in the *Low false alarm rate EDSCB*). As shown in Figure 2, automation with a hit rate of 86% and a high false alarm rate of 20% (*State-of-the-art EDSCB*) resulted in operators ignoring about one-half of the true automation alarms on difficult targets (bare explosives). Similar results have been reported by Hättenschwiler et al. (2018), using single-view EDSCB with on-screen alarm resolution. These results can be interpreted as a strong *cry-wolf effect*. Operators ignore automation alarms if they have experienced many false alarms from automation in the past (Breznitz, 1983), resulting from low compliance due to a low PPV.

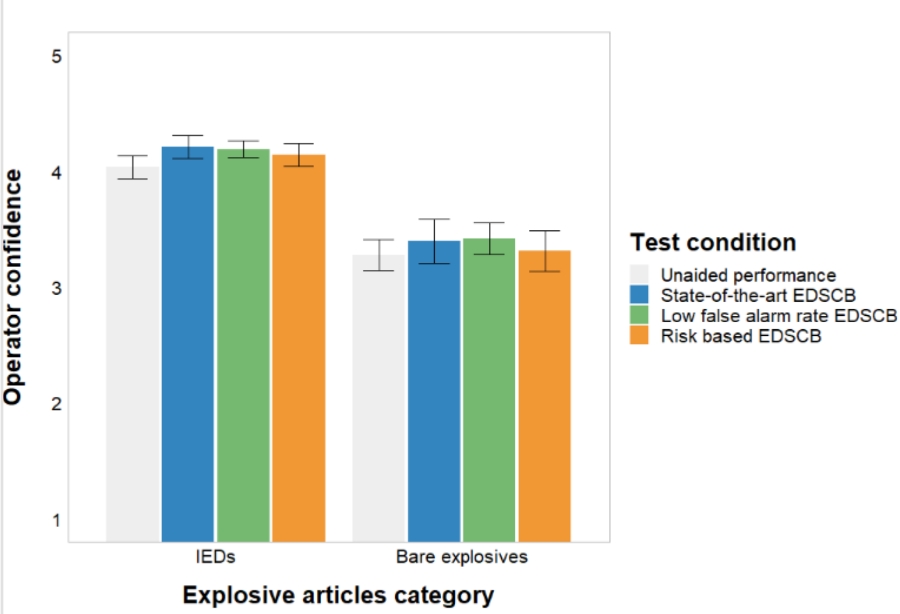


Figure 3: Mean operator confidence by test condition and explosive articles category. Error bars show ±1 standard error of the mean.

OPERATOR COMPLIANCE DEPENDS ON THE POSITIVE PREDICTIVE VALUE (PPV)

Figure 4 shows that screeners' compliance depended on the PPV of automation. Automation with a PPV of .5 (*Low false alarm rate EDSCB*) resulted in higher compliance than automation with a PPV of .3 (*State-of-the-art EDSCB* and *Risk based EDSCB*). This finding is consistent with previous research suggesting that operator compliance relates to the PPV of automation (e.g., Bartlett & McCarley, 2017, 2019; Getty et al., 1995; Meyer et al., 2014). Interestingly screeners complied at a similar ratio with the PPV as automation detected a target when an alarm was triggered. When automation had a PPV of .5 (*Low false alarm rate EDSCB*), screeners complied with automation alarms 48% of the time. When automation had a PPV of .3 (*State-of-the-art EDSCB* and *Risk based EDSCB*), screeners complied with automation alarms 28% and 32% of the time, respectively. This result has led to the *cry-wolf effect* shown in Figure 2. When alarmed images contained IEDs, screeners showed more compliance with automation alarms than when alarm images contained bare explosives. This could have been due to much higher unaided performance and higher confidence in detecting IEDs than bare explosives. It is reasonable to infer that screeners responded more often with target-present when automation alarmed

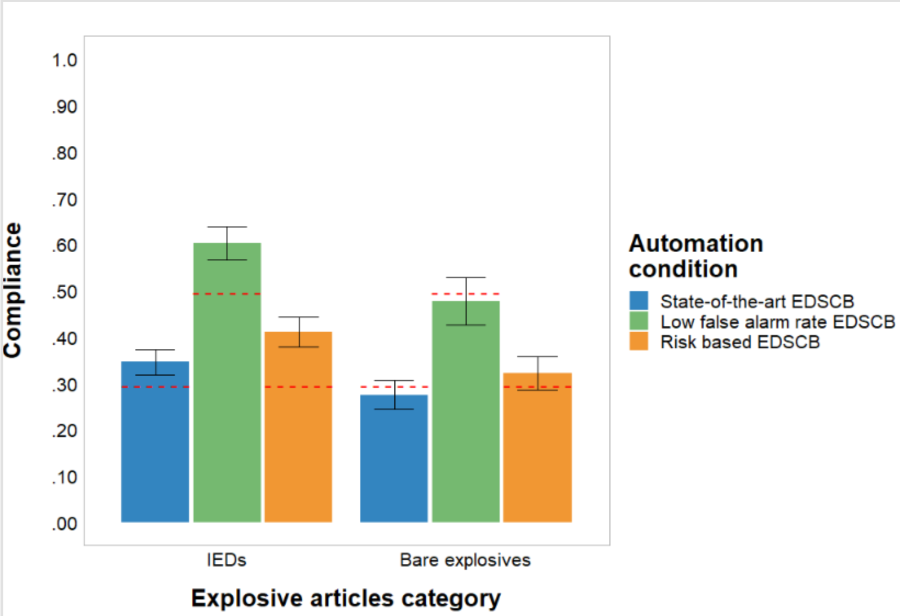


Figure 4: Mean operator compliance by explosives articles category and automation condition. Dashed red lines indicate the PPV of the automation. Error bars show ±1 standard error of the mean.

on an IED because they could detect IEDs well without automation and were more confident about doing this.

THE ADVANTAGE OF A CRY-WOLF EFFECT FOR EFFICIENCY

Using automation as operator support resulted in minor increases in human-machine system false alarm rates compared to unaided performance (see Figure 5). This was the case to a similar degree for all automation conditions. Interestingly, when the automation produced numerous false alarms (*State-of-the-art EDSCB* with an automation false alarm rate of 20%), screeners cleared most of them. This resulted in a similar human-machine false alarm rate in the other two automation conditions with automation false alarm rates of only 5% (*Low false alarm rate EDSCB* and *Risk based EDSCB*). This is good news because it shows that screeners can reduce the system false alarm rate when EDSCB is used as low-level automation. Screeners complied more with correct automation alarms (automation hits) than wrong automation alarms (automation false alarms). Therefore, we can rule out the possibility that screeners ignored a proportion of automation alarms at random.

PRACTICAL IMPLICATIONS AND RECOMMENDATIONS

Our results show that the human-automation interaction for EDSCB with on-screen alarm resolution is far from perfect. However, it should be noted that the performance of EDSCB is still improving rapidly and some of the problems shown in this article may also be solved in the future. Certain multi-view EDSCB with updated detection algorithms now achieve automation hit rates above 80% with automation false alarm rates below 10%, achieving even higher automation reliability than the EDSCB used in this study. Moreover, EDSCB based on computer tomography, which are increasingly used at airports, seem to achieve even lower automation false alarm rates, possibly below 5%, at least in the near future, according to airports and manufacturers. This extremely high automation reliability based on computer tomography could reduce the *cry-wolf effect* found in our study due to the lower false alarm rate compared to older multi-view EDSCB. However, we also found that whether screeners follow EDSCB alarms (operator compliance) or not depends on the PPV of automation which may be low even if the overall reliability of the automation is high. Accordingly, if the PPV is below .5, we recommend giving clear instructions to screeners to send all bags for which the EDSCB has triggered an alarm to secondary screening. Providing clear instructions helps to avoid the *cry-wolf effect* [22] and resolving all automation alarms in secondary screening with explosive trace detection or manual search appears to be operationally feasible for 5% of all passenger bags without requiring additional staff resources [45]. An alternative to clear instructions for screeners would be automated decision-making using EDSCB as high-level automation. That is, bags on which an EDSCB has alarmed are sent automatically to secondary screening. Regardless of whether EDSCB is used as low or high-level automation, screeners should still visually inspect X-ray images not only for the presence of guns and knives, but also for IEDs. The latter is crucial because EDSCB misses certain IEDs that can be detected only by visual inspection [47].

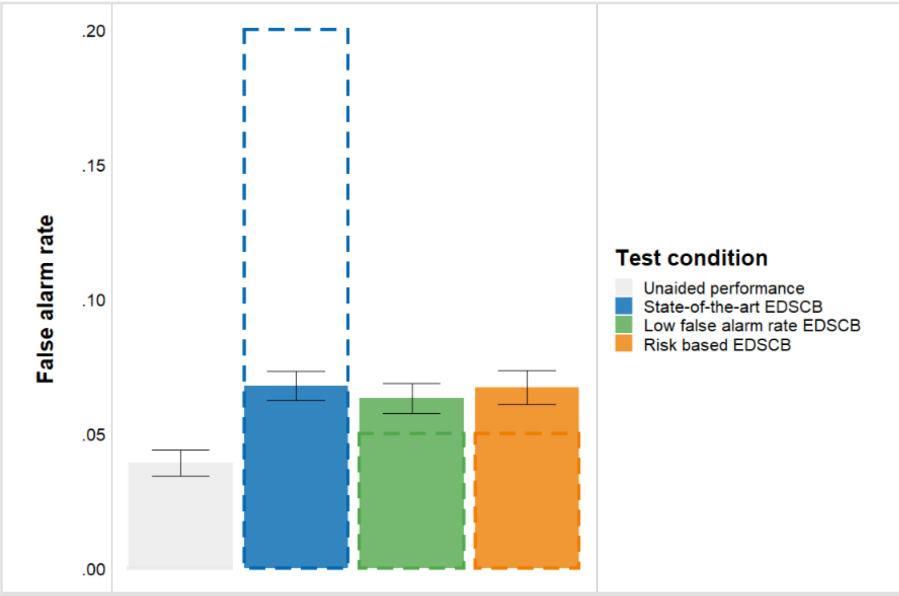


Figure 5 Human–machine system false alarm rate by test condition. Error bars show ±1 standard error of the mean. Dashed lines display the false alarm rates of automation by itself (automation false alarm rate).

REFERENCES

[1] R. Parasuraman, P. A. Hancock, and O. Olofinboba, "Alarm effectiveness in driver-centred collision-warning systems," *Ergonomics*, vol. 40, no. 3, pp. 390–399, Mar. 1997, doi: 10.1080/001401397188224.

[2] C. D. Wickens and S. R. Dixon, "The benefits of imperfect diagnostic automation: A synthesis of the literature," *Theor. Issues Ergon. Sci.*, vol. 8, no. 3, pp. 201–212, 2007, doi: 10.1080/14639220500370105.

[3] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," *IEEE Trans. Syst. Man,*

Cybern. - Part A Syst. Humans, vol. 30, no. 3, pp. 286–297, May 2000, doi: 10.1109/3468.844354.

[4] W. L. Sheridan, T. B. & Verplank, *Human and computer control of undersea teleoperators.* Cambridge, MA: MIT Man-Machine Systems Laboratory, Cambridge, 1978.

[5] M. Vagia, A. A. Transeth, and S. A. Fjerdingen, "A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed?," *Appl. Ergon.*, vol. 53, pp. 190–202, 2016, doi: 10.1016/j.apergo.2015.09.013.

[6] C. D. Wickens, A. Mavor, R. Parasuraman, and J. McGee, *The future of air traffic control: Human operators and automation.* Washington D.C.: National Academies Press, 1998.

[7] M. T. Dzindolet, L. G. Pierce, H. P. Beck, and L. A. Dawe, "The perceived utility of human and automated aids in a visual detection task," *Hum. Factors,* vol. 44, no. 1, pp. 79–94, 2002, doi: 10.1518/0018720024494856.

[8] D. Morrow, R. North, and C. D. Wickens, "Reducing and mitigating human error in medicine," *Rev. Hum. Factors Ergon.*, vol. 1, no. 1, pp. 254–296, 2005.

[9] D. A. Wiegmann, J. S. McCarley, A. F. Kramer, and C. D. Wickens, "Age and automation interact to influence performance of a simulated luggage screening task.," *Aviat. Space. Environ.* Med., vol. 77, no. 6, pp. 825–831, 2006.

[10] N. Hättenschwiler, Y. Sterchi, M. Mendes, and A. Schwaninger, "Automation in airport security X-ray screening of cabin baggage: Examining benefits and possible implementations of automated explosives detection," *Appl. Ergon.*, vol. 72, pp. 58–68, Oct. 2018, doi: 10.1016/j.apergo.2018.05.003.

[11] D. Huegli, S. Merks, and A. Schwaninger, "Automation reliability, human–machine system performance, and operator compliance: A study with airport security screeners supported by automated explosives detection systems for cabin baggage screening," *Appl. Ergon.*, vol. 86, no. April, 2020, doi: 10.1016/j.apergo.2020.103094.

[12] P. De Vries, C. Midden, and D. Bouwhuis, "The effects of errors on system trust, self-confidence, and the allocation of control in route planning," *Int. J. Hum. Comput. Stud.*, vol. 58, pp. 719–735, 2003, doi: 10.1016/S1071-5819(03)00039-9.

[13] S. R. Dixon, C. D. Wickens, and J. S. McCarley, "On the independence of compliance and reliance: Are automation false alarms worse than misses?," *Hum. Factors,* vol. 49, no. 4, pp. 564–72, 2007, doi: 10.1518/001872007X215656.

[14] P. Madhavan, D. Wiegmann, and F. Lacson, "Automation failures on tasks easily performed by operators undermine trust in automated aids," *Hum. Factors,* vol. 48, no. 2, pp. 241–256, 2006, doi: 10.1518/001872006777724408.

[15] S. R. Dixon and C. D. Wickens, "Automation reliability in unmanned aerial vehicle control: a reliance-compliance model of automation dependence in high workload.," *Hum. Factors*, vol. 48, no. 3, pp. 474–486, 2006, doi: 10.1518/001872006778606822.

[16] M. Maltz and D. Shinar, "New alternative methods of analyzing human behavior in cued target acquisition," *Hum. Factors*, vol. 45, no. 2, pp. 281–295, Jun. 2003, doi: 10.1518/hfes.45.2.281.27239.

[17] R. D. Sorkin and D. D. Woods, "Systems with human monitors: A signal detection analysis," *Human-Computer Interact.*, vol. 1, pp. 49–75, 1985, doi: 10.1207/s15327051hci0101.

[18] Breznitz S, *Cry-wolf: The psychology of false alarm.* Hillsdale, NJ: Lawrence Erlbaum, 1983.

[19] R. Parasuraman and C. D. Wickens, "Humans: Still vital after all these years of automation," *Hum. Factors*, vol. 50, no. 3, pp. 511–520, 2008, doi: 10.1518/001872008X312198.

[20] A. Zirk, R. Wiczorek, and D. Manzey, "Do we really need more stages? Comparing the effects of likelihood alarm systems and binary alarm systems," *Hum. Factors,* 2019, doi: https://doi.org/10.1177/0018720819852023.

[21] D. J. Getty, J. A. Swets, R. M. Pickett, and D. Gonthier, "System operator response to warnings of danger: A labaratory investigation of the effects of the predictive value of a warning on human response time," *J. Exp. Psychol.* Appl., vol. 1, no. 1, pp. 19–33, 1995.

[22] D. Manzey, N. Gerard, and R. Wiczorek, "Decision-making and response strategies in interaction with alarms: the impact of alarm reliability, availability of alarm validity information and workload," *Ergonomics*, vol. 57, no. 12, pp. 1833–1855, Dec. 2014, doi: 10.1080/00140139.2014.957732.

[23] J. Meyer and Y. Bitan, "Why better operators receive worse warnings," *Hum. Factors,* vol. 44, no. 3, pp. 343–353, 2002, doi: 10.1518/0018720024497754.

[24] M. L. Bartlett and J. S. McCarley, "Benchmarking aided decision making in a signal detection task," *Hum. Factors*, vol. 59, no. 6, pp. 881–900, 2017, doi: 10.1177/0018720817700258.

[25] M. L. Bartlett and J. S. McCarley, "No effect of cue format on automation dependence in an aided signal detection task," *Hum. Factors,* vol. 61, no. 2, pp. 169–190, Mar. 2019, doi: 10.1177/0018720818802961.

[26] J. Meyer, R. Wiczorek, and T. Günzler, "Measures of reliance and compliance in aided visual scanning," *Hum. Factors,* vol. 56, no. 5, pp. 840–849, 2014, doi: 10.1177/0018720813512865.

[27] S. Rice and J. S. McCarley, "Effects of response bias and judgment framing on operator use of an automated aid in a target detection task," *J. Exp. Psychol. Appl.*, vol. 17, no. 4, pp. 320–331, 2011, doi: 10.1037/a0024243.

[28] D. A. Wiegmann, A. Rich, and H. Zhang, "Automated diagnostic aids: The effects of aid reliability on users' trust and reliance," *Theor. Issues Ergon. Sci.*, vol. 2, no. 4, pp. 352–367, 2001, doi: 10.1080/14639220110110306.

[29] T. Halbherr, A. Schwaninger, G. R. Budgell, and A. Wales, "Airport security screener competency: A cross-sectional and longitudinal analysis," *Int. J. Aviat. Psychol.*, vol. 23, no. 2, pp. 113–129, Apr. 2013, doi: 10.1080/10508414.2011.582455.

[30] S. M. Koller, D. Hardmeier, S. Michel, and A. Schwaninger, "Investigating training, transfer and viewpoint effects resulting from recurrent CBT of X-Ray image interpretation," *J. Transp. Secur.*, vol. 1, no. 2, pp. 81–106, 2008, doi: 10.1007/s12198-007-0006-4.

[31] D. Schuster, J. Rivera, B. C. Sellers, S. M. Fiore, and F. Jentsch, "Perceptual training for visual search," *Ergonomics*, vol. 56, no. 7, pp. 1101–1115, 2013, doi: 10.1080/00140139.2013.790481.

[32] A. Schwaninger and F. Hofer, "Evaluation of CBT for increasing threat detection performance in X-ray screening.," in *The Internet Society 2004, Advances in Learning, Commerce and Security*, K. Morgan and M. J. Spector, Eds. Ashurst, England: WIT Press, 2004, pp. 147–156.

[33] S. Turner, *Terrorist explosive sourcebook: countering terrorist use of improvised explosive devices.* Paladin Press, Boulder, CO, 1994.

[34] K. Wells and D. A. Bradley, "A review of X-ray explosives detection techniques for checked baggage," *Appl. Radiat. Isot.*, vol. 70, no. 8, pp. 1729–1746, 2012, doi: 10.1016/j.apradiso.2012.01.011.

[35] S. M. Koller, C. G. Drury, and A. Schwaninger, "Change of search time and non-search time in X-ray baggage screening due to training.," *Ergonomics*, vol. 52, no. 6, pp. 644–56, Jun. 2009, doi: 10.1080/00140130802526935.

[36] T. L. Jones, *Court security: A guide for post 9-11 environments.* Springfield, Illinois: Charles Thomas Publisher, 2003.

[37] K G. Zentai, "X-ray imaging for homeland security," 2008 I*EEE Int. Work. Imaging Syst. Tech.*, pp. 1–6, 2008, doi: 10.1109/IST.2008.4659929.

[38] F. Hofer and A. Schwaninger, "Using threat image projection data for assessing individual screener performance," *WIT Trans. Built Environ.*, vol. 82, pp. 117–126, May 2005, doi: 10.2495/SAFE050411.

[39] R. F. Meuter and P. F. Lacherez, "When and why threats go undetected: Impacts of event rate and shift length on threat detection accuracy during airport baggage screening," *Hum. Factors,* vol. 58, no. 2, pp. 218–228, 2016, doi: 10.1177/0018720815616306.

[40] A. Schwaninger, D. Hardmeier, J. Riegelnig, and M. Martin, "Use it and still lose it ? The influence of age and job experience on detection performance in X-ray screening," *GeroPsych* (Bern)., vol. 23, no. 3, pp. 169–175, 2010, doi: 10.1024/1662-9647/a000020.

[41] J. Skorupski and P. Uchroński, "A human being as a part of the security control system at the airport," *Procedia Eng.*, vol. 134, pp. 291–300, 2016, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877705816000138.

[42] A. Schwaninger, "Why do airport security screeners sometimes fail in covert tests?," in *Proceedings of the 43rd IEEE International Carnahan Conference on Security Technology*, 2009, no. October 5-8, pp. 41–45, doi: 10.1109/CCST.2009.5335568.

[43] O. Wetter, D. Hardmeier, and F. Hofer, "Covert testing at airports: Exploring methodology and results," 2008, doi: 10.1109/CCST.2008.4751328.

[44] Commission Implementing Regulation (EU), "Laying down detailed measures for the implementation of the common basic standards on aviation security 2015/1998 of 5 November 2015." Official Journal of the European Union, 2015.

[45] Y. Sterchi and A. Schwaninger, "A first simulation on optimizing EDS for cabin baggage screening regarding throughput," in *Proceedings of the 49th IEEE International Carnahan Conference on Security Technology,* Taipei Taiwan, September 21–-24, 2015, pp. 55–60, doi: 10.1109/CCST.2015.7389657.

[46] P. Baum, *Violence in the skies: a history of aircraft hijacking and bombing.* Chichester, England: Summersdale Publishers, 2016.

[47] J. Howell, "The modern IED : Design and trends," *Aviation Security International Magazine,* pp. 1–15, 2017.

INTERNATIONAL CARGO SCREENING STUDY

Text: Slavtcho Groshev

This article presents a web-based international cargo screening study conducted in cooperation with the International Air Transport Association (IATA) and id3 Technologies in 2020 with participating entities located across multiple continents. A central focus of the study was illustrating benefits of computer-based training (CBT) for cargo screening along with potential improvements.

STUDY BACKGROUND, OBJECTIVES, AND DESIGN

The IATA Cargo Security and Facilitation Forum at the end of October 2019 featured, among other topics, a presentation by CASRA that facilitated discussions and engagement on the importance of human factors in cargo screening [1]. Industry interest was expressed in examining screener performance, the use of computer-based training (CBT) and the value-added it creates. Additionally, the role of security culture was to be considered, with the 40th Session of the International Civil Aviation Organization (ICAO) Assembly designating 2020 as the Year of Security Culture (YOSC) [2].

An international study was subsequently conducted in 2020 with the aim of examining the following aspects:

- Improvement of image interpretation performance through individually adaptive CBT with X-Ray Tutor (XRT)

INFOBOX: IATA AND BIOSEAL

The International Civil Aviation Organization (ICAO) Aviation Security (AVSEC) Panel and ICAO Council encourage states and all stakeholders to rapidly review and update the national requirements and syllabus of existing security awareness training programs to include incident reporting awareness and processes. In order to assist with this rapid implementtion, IATA launched the *See It Report It* initiative. The initiative includes videos on awareness and reporting, both available in 13 languages. For more details see [here].

BioSeal is a technology co-developed by id3 Technologies. It is a robust and economical solution for both combating the fraudulent forgery of all types of documents and managing the identity of people through biometrics. It ensures the authenticity and integrity of non-electronic documents in a comparatively inexpensive, but highly secure manner using asymmetric cryptography. BioSeal is a secure and General Data protection Regulation (EU GDPR) compliant solution for biometric authentication. For more detils see [here].

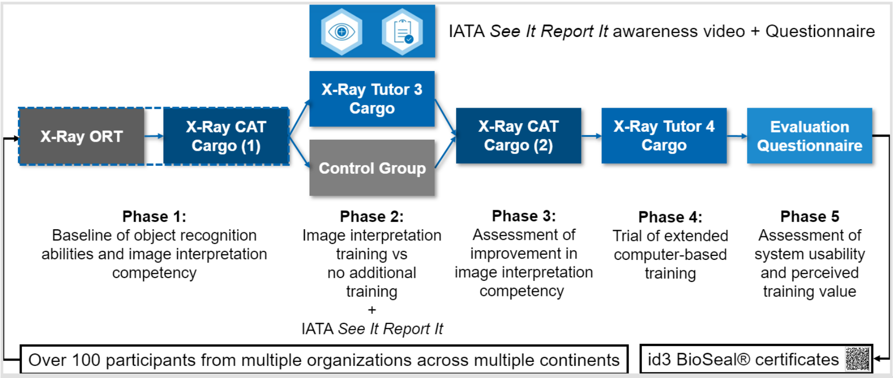


Figure 1: Study design

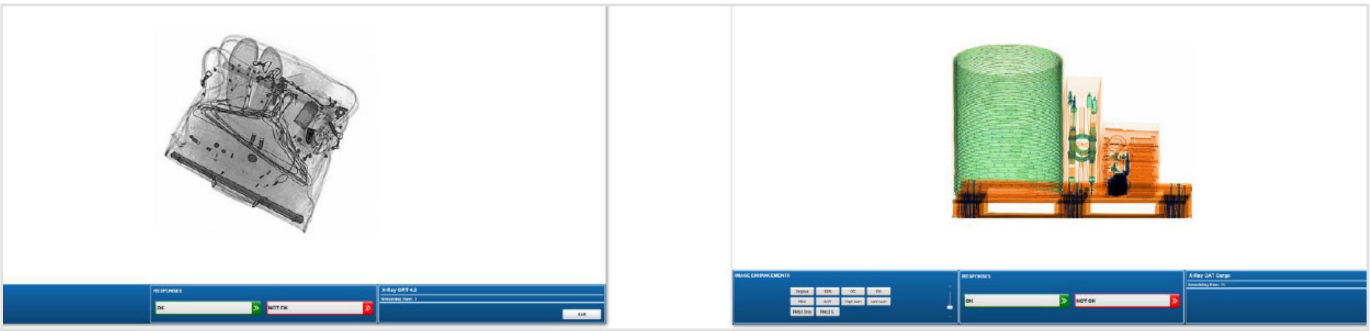


Figure 2: Phase 1 - X-Ray ORT (left) and X-Ray CAT Cargo (right)

- Usefulness of well-established X-ray operator (screener) selection tool X-Ray Object Recognition Test (X-Ray ORT) for the domain of cargo screening
- Potential enhancements of CBT for cargo screening through incorporation of aspects such as handling of multiple X-ray views, shielding, and air waybills (AWB)
- Reception of IATA *See It Report It* awareness video

As noted by the Guidance for ground handling during and post COVID-19 by IATA, the pandemic has led to enhanced importance of remote training [3], so the study can serve as an applied example. A naturally emerging consideration within this context is identity verification for remote training certificates, and so id3 Technologies was supporting the study with secure digital certificates for participants to demonstrate applicability of their technology for this use case.

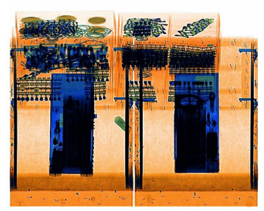


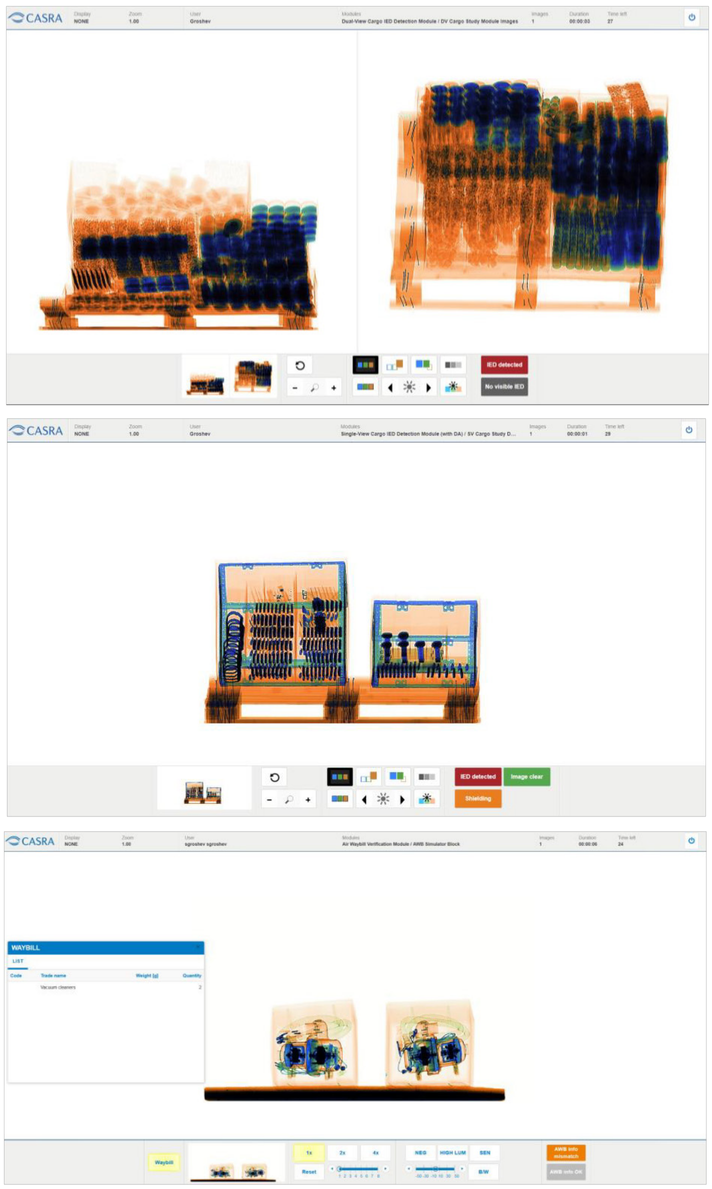
Figure 3: Phase 2 - XRT3 Cargo

More than 100 participants from multiple organizations across multiple continents were involved in a five-phase study conducted fully online (Figure 1). The study phases will be explained in detail in the following sections.

STUDY PHASES

Phase 1 of the study served to establish a baseline of object recognition abilities and image interpretation competency. The scientifically validated instruments X-Ray ORT and X-Ray Competency Assessment Test (X-Ray CAT) Cargo were used (Figure 2). The X-Ray ORT assesses the ability to deal with viewpoint, superposition, and bag complexity (so-called image-based factors that have been scientifically demonstrated to strongly influence X-ray image difficulty) and features 256 images shown for four seconds each, half of which contain either a knife or a handgun. The test is geared towards pre-employment selection and does not require previous knowledge of X-ray image interpretation [4,5]. That is why the pictures are in black and white. The X-Ray CAT Cargo assesses the ability to interpret cargo images and identify improvised explosive devices (IEDs). It features 160 images shown for 15 seconds each, half of which contain an IED [6,7].

Phase 2 of the study featured a random assignment of participants to the training (experimental) or the no training (control) group. The experimental group trained with the individually adaptive CBT X-Ray Tutor 3 (XRT3), a scientifically validated tool



Phase 4 - XRT4 dual-view (top), shielding (middle), and AWB verification (bottom) modules for image interpretation training (Figure 3) [7]. Images with and without IEDs were presented based on level progress and individual previous detection «history» of trainees. An average of 10.9 hours were trained, however with large differences between participants (standard deviation of 3.6 hours).

In parallel, participants had access to a short security awareness video from the IATA *See It Report It* initiative and were asked to answer three questions after watching it as a concise assessment of its reception and prospective contribution to security culture.

Phase 3 of the study served to assess the improvement in X-ray image interpretation competency. The expected results upon repetition of the X-Ray CAT Cargo were an improvement in X-Ray CAT performance for the group that trained and no improvement for the group that did not train. Further, we expected a positive relationship between the X-Ray ORT and X-Ray CAT performance independently of training, since higher detection performance can be presumed for individuals with object recognition abilities independent of training.

Phase 4 of the study featured a trial of three short modules addressing additional training aspects, namely dual-view imaging, shielding, and AWB verification within XRT4 (Figure 4) [8]. The dual-view module presented freight from two perspectives, asking the participant to decide if they could identify a hidden IED. The shielding module included the decision option of sending the cargo for further inspection due to shielding. Finally, the AWB module required participants to verify if what was declared on the AWB matched the X-ray image. All three modules required participants to indicate the confidence of their decision on a five-point scale. The expected results were for the second X-Ray CAT performance to correlate positively with the dual-view module confidence weighted performance score (CWPS, for calculation see results section), the X-Ray ORT performance to correlate positively with the shielding module CWPS, and screening experience to correlate positively with AWB module CWPS.

Phase 5 involved an assessment of the usability and perceived training value of the XRT4 modules as well as

	X-Ray CAT 2 A'	Experience	Shielding Module CWPS	Dual-View Module CWPS	AWB Module CWPS
X-Ray ORT A'	0.70***	0.22*	0.34**	0.36**	0.29*
X-Ray CAT 2 A'		0.24**	0.47***	0.50***	0.089
Experience			0.15	0.14	0.18
Shielding Module CWPS				0.58***	0.19
Dual-View Module CWPS					0.24*

*Figure 5: Correlations between variables with statistical significance (*** : p < .001, ** : p < .01, * : p < .05), CWPS calculation pseudo code for confidence measured on five-point scale: IF TargetButton = ClickedButton THEN Score = 1 ELSE Score = -1 >> CWPS = (Score * Confidence + 5) / (5 * 2)*

CBT for cargo screening in a broader sense. Questions were taken from well-established measurement scales, namely the System Usability Scale (SUS) [9] the Training Evaluation Inventory (TEI) [10], and extended with specific questions of relevance for CBT for cargo screening.

STUDY RESULTS

The results of the study revealed that the X-Ray ORT performance correlated strongly and highly significantly with the X-Ray CAT performance and weakly but highly significantly with the shielding module CWPS, confirming the expectation that the X-Ray ORT can predict future performance in image interpretation for cargo screening. The detection performance in the second X-Ray CAT Cargo measurement correlated moderately and highly significantly with the dual-view module CWPS, confirming the expectation that screeners who perform well in single-view image interpretation are likely to perform well in dual-view image interpretation. Screening experience did not correlate significantly with the AWB module CWPS. This could have been due to the rather simple AWBs used. The main findings were robust within an analysis of variance including age and group assignment.

The extended CBT within XRT4 was well received overall, with a majority of respondents indicating a subjective expansion

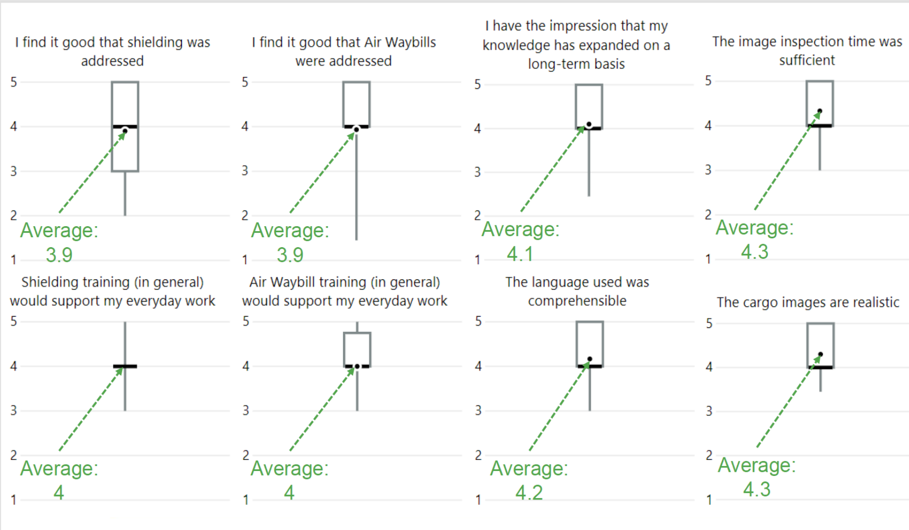


Figure 6: Training evaluation questionnaire results on a five point scale (1 = do not agree at all, 5 = agree fully)

of knowledge and a desire for additional shielding and AWB training (Figure 5). Potential for improvement was, however, also highlighted for the specific shielding and AWB training modules. Some screeners indicated for example that some AWBs were very difficult due to similarity between the appearance of the falsely declared and the actually presented goods (requiring good visual memory) and others noted different rules for classification of shielding being applicable in their work environment.

The IATA *See It Report It* awareness video was very well received overall, with almost all respondents indicating agreement or full agreement with a recommendation for including the video in their work environment training and awareness campaigns (Figure 6). Wishes for a specific version for cargo screening were also expressed.

CONCLUSION

The study demonstrated that substantial improvements in cargo screener performance can be achieved (average of +8% hits and -6% false alarms) through individually adaptive online CBT (in particular using XRT) after an average of around eleven hours of training already. The screener selection tool X-Ray ORT was further shown to be useful for cargo screening as well due to its power to predict image interpretation performance of screeners independently of training. The value of screener feedback for further development and customizing of features of CBT for cargo screening was also highlighted. The very good reception of the IATA *See It Report It* online awareness video showed that it is a promising tool for supporting security culture. Last but not least, potential for international online performance benchmarking in cargo X-ray screening as well as using biometric certificates for online training and assessment was demonstrated.

We would like to thank all participating entities, IATA, and id3 Technologies for their valuable roles in this effort.

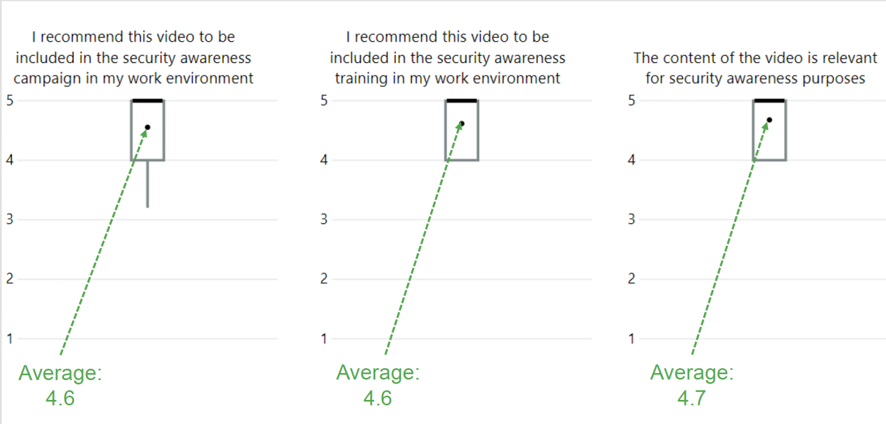


Figure 7: IATA See It Report It awareness video evaluation questionnaire results (1 = do not agree at all, 5 = agree fully)

REFERENCES

[1] IATA. (2019). *IATA Cargo Security & Facilitation Forum*, 29 – 30 October 2019, Amsterdam, Netherlands - Day 2. [PDF]

[2] ICAO. (2021). *Year of Security Culture 2021.* [Website]

[3] IATA. (2021). *Guidance for ground handling during and post COVID-19.* [PDF]

[4] CASRA. (2020). X-Ray ORT – *X-Ray Object Recognition Test.* [Website]

[5] 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. In *Proceedings 39th Annual 2005 International Carnahan Conference on Security Technology*, 189-192. IEEE.

[6] CASRA. (2020). X-Ray CAT – *X-Ray Competency Assessment Test.* [Website]

[7] Michel, S., Mendes, M., de Ruiter, J. C., Koomen, G. C., & Schwaninger, A. (2014). Increasing X-ray image interpretation competency of cargo security screeners. *International journal of industrial ergonomics*, 44(4), 551-560.

[8] CASRA. (2020). XRT4 – *X-Ray Tutor 4.* [Website]

[9] Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human–Computer Interaction*, 24(6), 574-594.

[10] Ritzmann, S., Hagemann, V., & Kluge, A. (2014). The Training Evaluation Inventory (TEI)-evaluation of training design and measurement of training outcomes for predicting training success. *Vocations and Learning*, 7(1), 41-73.

Science & Security – CASRA Newsletter
ISSN 1664-5715

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

Editorial Manager:
Mahé Becker

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