SCIENCE & SECURITY Issue 26

CASRA NEWSLETTER

It is time to publish our second newsletter in 2021!

Since 2012, we have been publishing newsletters in which we present results from our research, address trends, and provide information for security practitioners.

In this issue, we present a research and development project on comparing airport security checkpoints regarding effectiveness and efficiency. The second article presents a project conducted in collaboration with the International Air Transport Association (IATA) on assessing X-ray image interpretation competency of X-ray hold baggage screeners..

We hope you enjoy reading these new articles and 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.

With best wishes,

Dr. Diana Hardmeier
Director

Prof. Dr. Adrian Schwaninger
Chairman

TOPICS IN THIS ISSUE:

RESEARCH PUT ACROSS

COMPARING THE EFFECTIVENESS AND EFFICIENCY OF CHECKPOINTS

As part of a research and development project, CASRA has developed methods and tools to assist with evaluating the different performance aspects of airport security checkpoints: Security, efficiency, passenger experience, and a work environment that allows for motivated security officers. This article presents one of these tools, which can be used to compare the overall security provided by different checkpoint technologies and configurations.

SECURITY IN PRACTICE

INTERNATIONAL BENCHMARKING FOR X-RAY HOLD BAGGAGE SCREENING

This article presents the International Air Transport Association (IATA) X-ray hold baggage screening assessment along with the benchmarking methodology and tool. They were developed and piloted in 2021 within the IATA Innovation in Security – Processes and Technology Workstream#3 led by CASRA. The methodology and tool were developed based on industry contributions and leverage more than 15 years of applied research in the area of human factors in aviation security screening.

COMPARING THE EFFECTIVENESS AND EFFICIENCY OF CHECKPOINTS

Text: Yanik Sterchi

As part of a research and development project, CASRA has developed methods and tools to assist with evaluating the different performance aspects of airport security checkpoints: Security, efficiency, passenger experience, and a work environment that allows for motivated security officers. This article presents one of these tools, which can be used to compare the overall security provided by different checkpoint technologies and configurations.

CHECKPOINT PERFORMANCE

Successfully designing and running an airport security checkpoint is a challenge. It has to meet high security standards, provide sufficient throughput at acceptable cost, offer a positive experience to passengers, and should also provide a friendly workplace for the security personnel.

To find the optimal balance between these different aspects, you need information on how well your checkpoint performs on them. Ideally, you also get information from other airports on how well their checkpoints perform. In the research project "4CAST", CASRA explores and develops different methods to evaluate checkpoint performance, including a tool to estimate the overall level of security provided by a checkpoint, a survey designed to compare the efficiency of different checkpoints, video observation to measure process times, 3D simulation modelling of checkpoints, and an employee survey measuring how different aspects of the screeners' work affect their motivation, satisfaction, and turnover-intention. In this article, we present the tool for the estimation of the overall security provided by a checkpoint.

CHECKPOINT EFFECTIVENESS

Today, there are many options when choosing the equipment for an airport security checkpoint. For the screening of cabin baggage, they range from single view X-ray machines to state-of-the-art computed tomography (CT) machines with automated explosives detection (EDS); for passenger screening one has to decide between a variety of walk-through metal detectors (WTMDs) and millimeter wave security scanners. Of all the options that meet the current regulatory standards, not all provide the same level of security – certainly not for all kinds of threats. For example, a checkpoint with security scanners is much more likely to detect a ceramic knife hidden on the body compared to a checkpoint with WTMDs. But how much more likely? To answer this question, we would have to know how likely a ceramic knife triggers an alarm of a security scanner and of a WTMD, but also how likely the respective alarm resolution processes detect the threat and whether it might be detected by other screening processes performed at the checkpoint. If we assume that the security scanner alarms with a probability of 90% and alarms are resolved with a targeted pat-down that again finds the threat with a probability of 90%, then – assuming independent probabilities and that none of the other screening processes might detect the threat – the overall probability of the ceramic knife being found would be 81%. In comparison, a WTMD would be less likely to trigger an alarm and alarms would be resolved with full-body pat-downs instead of targeted pat-downs. Also, at a checkpoint using WTMDs, some passengers might be randomly selected for pat-downs. Again, when the respective probabilities are known, an overall probability of detection can be estimated

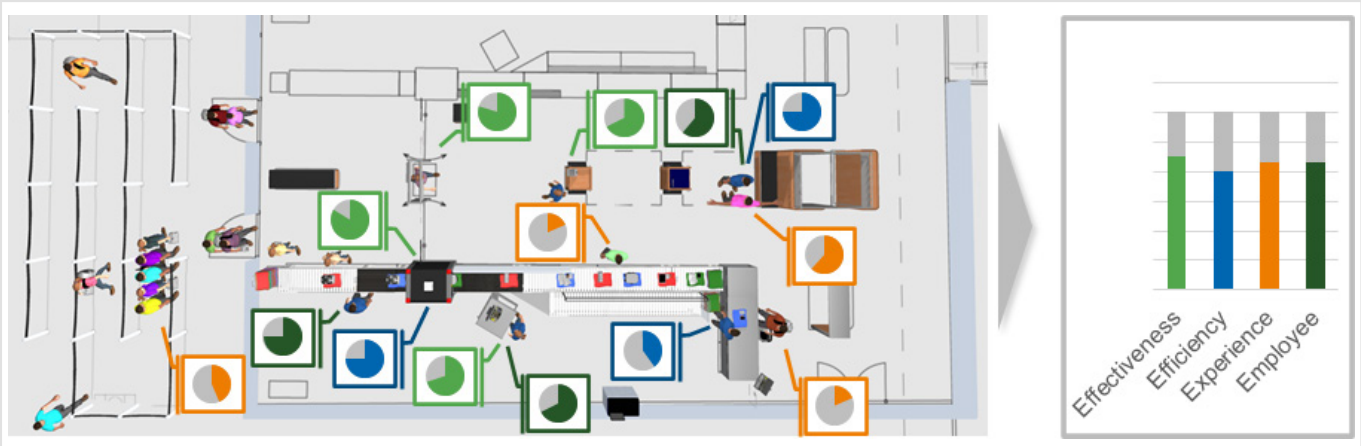

Figure 1: A myriad of components, procedures, and interactions affect the different performance aspects of an airport security checkpoint as a whole.

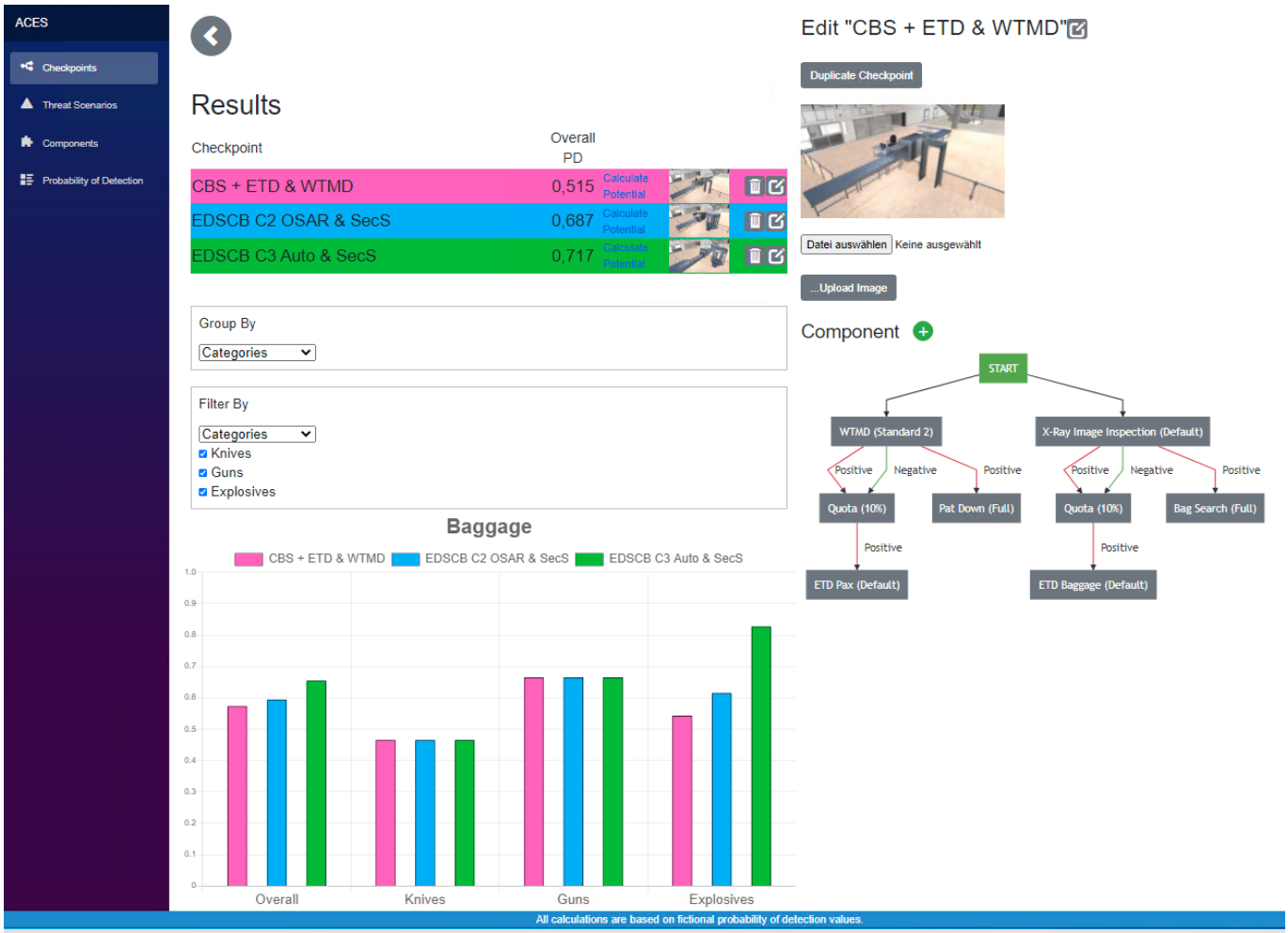

Figure 2: Screenshot of the ACES tool. On the left side, overall checkpoint effectiveness and effectiveness for single threat categories are shown. The right side shows the configuration of a selected checkpoint.

and one could compare the effectiveness of the two checkpoints in detecting ceramic knives.

AIRPORT CHECKPOINT EFFECTIVENESS

As illustrated in the previous paragraph, it is – in principle – possible to compare the overall effectiveness of different checkpoint configurations. But there are some caveats. A first challenge is to know how likely the components and screening processes trigger an alarm in presence of certain threats. For technical components, this can be derived from the detection probabilities that are minimally required by regulation or from data gathered in component testing for their certification. For human-machine systems, data from studies and covert tests can provide estimates (see, for example, our previous newsletter on a study investigating performance of screeners assisted by EDS).

The second challenge is to efficiently manage all the relevant information. You would likely want to look at many different kinds of threats, you would want to know how likely they are detected if hidden in the baggage and if hidden on the body. For some of these threats, many different components and screening processes have to be considered: Explosives hidden in baggage might be detected through a manual search resolving an alarm of the X-ray image inspection, through an ETD (explosives trace detection) resolving an alarm of the EDS, through a random ETD performed on the bag, or maybe even through a random ETD performed on the passenger. In addition to managing many threat scenarios and components, you might want to compare several different checkpoint configurations and explore how small changes to one or multiple of these check-points affect their overall effectiveness.

As part of the research project "4CAST" funded by the FOCA, CASRA has developed a tool that assists with evaluating the overall effectiveness of checkpoints. The "Airport Security Check-point Effectiveness Simulator" (ACES; see Figure 2 for a screenshot) allows the user to define the checkpoints that should be compared and the threat scenarios for which the effectiveness should be calculated. As a third input, ACES requires the detection probabilities for each component and threat scenario. It then calculates the detection probability of each checkpoint on different levels of aggregation (per threat scenario, per threat category, and over all threat categories combined). The results are displayed as graphs or can be exported to Microsoft Excel.

USE AND FURTHER DEVELOPMENT

Better understanding how checkpoints differ in their overall effectiveness against various threats can also help guiding future regulation. Therefore, Dutch and German regulators are exploring the capabilities of ACES to estimate the overall effectiveness of the checkpoint configurations frequently used in their countries and to compare different regulatory standards. Their valuable feedback has contributed to the continuous development of ACES. Within the resources of the research project, CASRA can provide the tool for free to you and continue its development based on your feedback.

To design the optimal checkpoint, a better understanding of the effectiveness of different checkpoints is certainly key, but not sufficient. Together with several airports, CASRA therefore assesses how efficient common checkpoint configurations are, when and why passengers experience the security check positively, and what fosters the motivation and satisfaction of the security personnel. Taking all of these aspects into account, checkpoints will hopefully become more secure, fast and affordable, as well as more convenient for both passengers and security officers.

INTERNATIONAL BENCHMARKING FOR X-RAY HOLD BAGGAGE SCREENING

Text: Slavtcho Groshev

This article presents the International Air Transport Association (IATA) X-ray hold baggage screening assessment along with the benchmarking methodology and tool. They were developed and piloted in 2021 within the IATA Innovation in Security – Processes and Technology Workstream#3 led by CASRA. The methodology and tool were developed based on industry contributions and leverage more than 15 years of applied research in the area of human factors in aviation security screening.

THE IMPORTANCE OF HUMAN FACTORS IN SECURITY SCREENING

Resolution 40-11 was adopted by the 40th Session of the International Civil Aviation Organization (ICAO), designating 2020 as the Year of Security Culture (YOSC). With the COVID-19 pandemic severely impacting aviation in 2020, the ICAO Council decided to extend the YOSC to 2021. Additionally, 2021 marks 20 years since 9/11 - the worst acts of unlawful interference in the history of aviation. It is thus imperative to understand the threat to aviation and promote best practices in security throughout all aviation operations [1].

Regulators have noted that COVID-19 has meant that the aviation industry has had to take difficult decisions regarding staffing. In many jurisdictions, this has involved a temporary scaling down of aviation operations, often involving furlough or considerable periods of staff not working. However, there is no reason to believe that COVID-19 has reduced the attractiveness of aviation as a terrorist target [2].

Several activities were initiated by IATA, its members and partner organizations to leverage the emphasis on security culture and improve security performance in critical areas, starting from front-line personnel and those in specialized security functions (e.g. operators of X-ray screening equipment) up to the managerial level. Human performance in security screening remains a key element of protective measures. It is therefore essential it can be measured using fair, transparent and objective set of criteria [4].

The IATA Innovation in Security – Processes and Technology Workstream#3 initiated in 2020 and led by CASRA focused on this aspect.

INTERNATIONAL BENCHMARKING INITIATIVE CONTEXT AND OVERVIEW

The IATA Innovation in Security – Processes and Technology Workstream#3 focused on developing recognized best practices and standards to ensure human resources, and in particular screening equipment operators (screeners), can deliver good quality screening consistently. A core set of performance criteria for screeners was endorsed based on industry input and research results, to be followed by the development of an assessment tool (image library + assessment protocol) allowing to verify performance level achievement in a valid, reliable and standardized manner [3].

Input received from airports combined with CASRA in-house expertise were leveraged to create an image library containing images without prohibited articles and images with improvised explosive / incendiary devices. This image library was to be used following an assessment protocol independent of the particular software used, resulting in an assessment tool (Figure 1).

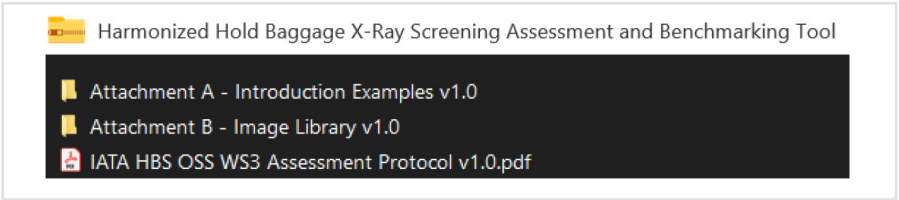

Figure 1: Harmonized hold baggage X-ray screening assessment and benchmarking tool structure.

INTERNATIONAL BENCHMARKING INITIATIVE PILOT

A pilot was conducted online in March 2021 using an implementation of the assessment tool within X-Ray Tutor 4 (XRT4), the main web-based training and assessment platform provided by CASRA. International involvement of industry across Europe, the Middle East, New Zealand, South Africa, South America and more than 100 screeners allowed gathering data and refining the tool. Figure 2 shows the implementation in XRT4, where participants were able to securely login online and get access to a module built in accordance with the abovementioned assessment protocol.

INTERNATIONAL BENCHMARKING INITIATIVE RESULTS ANALYSIS

Signal detection theory provides performance metrics based on four basic outcomes in a decision setting (Figure 3) [4]. Not only is the hit rate (i.e. the proportion of rejected/alarmed bags containing threats) important, but so is the false alarm rate (i.e. the proportion of rejected/alarmed harmless bags). This reflects the

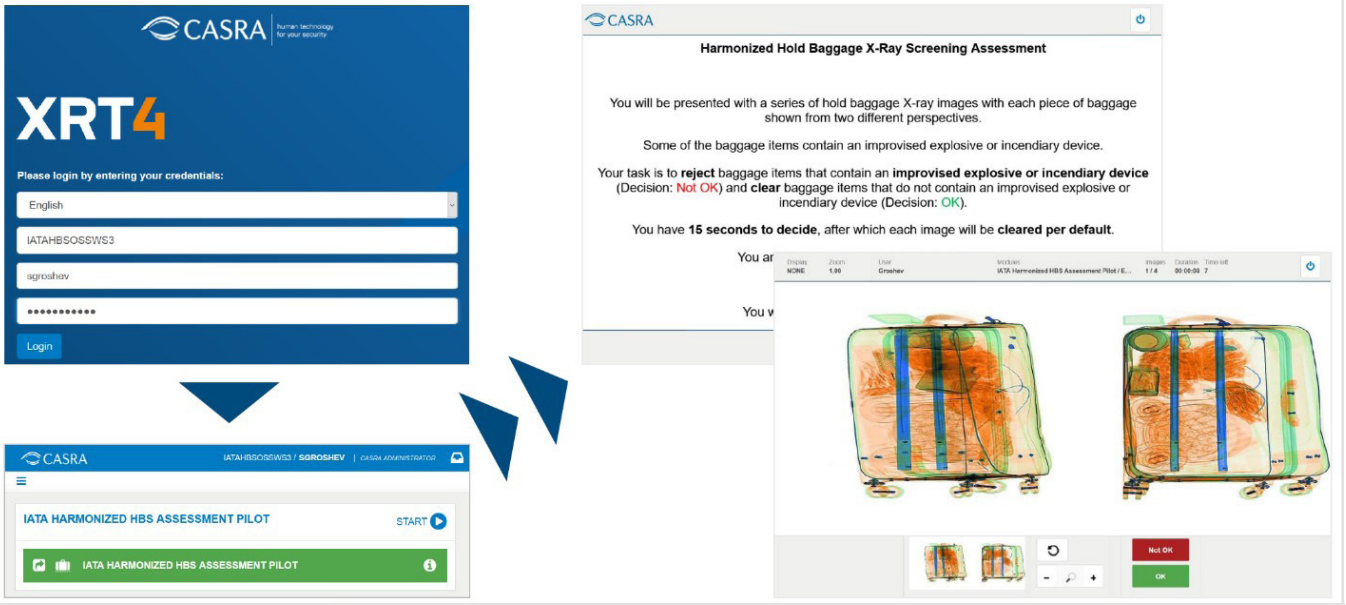

Figure 2: Harmonized hold baggage X-ray screening assessment and benchmarking tool pilot implementation.

understanding of detection performance as the ability not only to detect prohibited items but also to discriminate between prohibited and allowed items, and it additionally greatly facilitates performance comparison. Overall performance metrics such as A′ combine hit rate and false alarm rate into one performance measure, with A′ = 0.5 meaning chance performance and A′ = 1.0 perfect performance [5].

The basic assumption for the analysis was that the test was composed of items (images) measuring a single underlying ability, namely the ability to interpret X-ray images of hold baggage correctly so as to distinguish images containing improvised explosive or incendiary devices. The quality of the test as a whole was thus assessed by estimating its internal consistency (i.e. to what extent test items yield similar results), aiming for split-half reliability above 0.75 and Cronbach's alpha above 0.85 respectively. [6] To achieve the desired high internal consistency, an acceptable range was defined for the average correct responses per item, assuming a representative pilot sample. Items were then selected that had achieved both item-scale correlation and discrimination coefficients of at least 0.2. [4] The resulting test was to take under 20 minutes, approximating short image interpretation work sessions, all while keeping excellent internal consistency (Cronbach's alpha above 0.9). The target prevalence was set at 1:2 as a compromise between statistical power (strengthened by target prevalence close to 1:1) and approximation of operational conditions using threat image projection (substantially below 1:1 for operational reasons) [7].

Previous aviation security reference levels work had provided detection performance level reference values, sug-

$$A' = 0.5 + \frac{(HR - FAR)(1 + HR - FAR)}{4HR(1 - FAR)} \mid HR \geq FAR$$

Stimulus given \ Decision	Target absent / No signal / Bag is harmless	Target present / Signal / Bag requires secondary search
Target absent / Noise / No threat present	Correct rejection	False alarm
Target present / Signal plus noise / Threat present	Miss	Hit

Example:

40 clear bags cleared	20 clear bags alarmed *FAR* = 20 / (40 + 20)
10 threat bags cleared	20 threat bags alarmed *HR* = 20 / (10 + 20)

Figure 3: Four basic outcomes in a decision setting based on signal detection theory (left), example calculation of hit rate (HR) and false alarm rate (FAR) (bottom right) and calculation of A′ (top right).

gesting a realistic mean expected detection performance. A combination of a normative and criterion-referenced pass mark definition approach was then used, defining what proportion of screeners should reach specified performance criteria in terms of hit rate, false alarm rate and overall performance (A') endorsed by Workstream#3 stakeholders [9]. The performance of participants in the pilot could subsequently be matched against these (increasingly challenging) performance criteria corresponding to three performance levels (C, B, A). Performance lower than level C was categorized as Failed. The resulting performance distribution is shown in Figure 4, with participants being categorized based on the highest performance level criteria they had met. The height of the vertical bars corresponds to the proportion of participants in the pilot meeting the respective level criteria, with roughly half of participants reaching level A and about one third failing to reach level C.

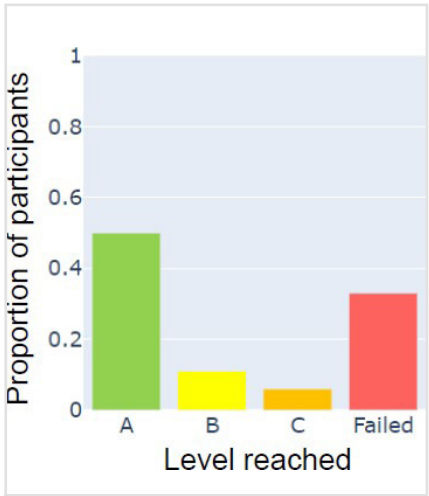

Figure 4: International benchmarking pilot performance distribution.

INTERNATIONAL BENCHMARKING INITIATIVE ACHIEVEMENTS

Many Workstream#3 stakeholders expressed support of the concept, methodology and outcomes of the pilot study and found the development valuable. It was highlighted that its applications may reach beyond the current initiative, as it could be generally helpful in informing strategies, investments, and professional development plans (e.g. in the scope of risk assessment processes and benchmarking). It was noted that a methodology transfer from this pilot, focused on hold baggage screening, is possible to screening of cargo and for wildlife, contraband and dangerous goods. Only limited reservations were expressed by Workstream#3 stakeholders, mainly seen as areas for consideration in future developments. In particular, a use of 3D images was noted as desirable and curiosity expressed regarding the expected impact on results of increasing sample size. All participants were encouraged to volunteer for further testing to increase data sample size allowing more detailed analysis and presentation of updated results (based on 1000 participants) by autumn 2021. Within this period, CASRA offered the test for no charge as a global benchmarking tool and support for organizations in evaluation of their screeners, should they have not been able to train/work to the same degree as before the pandemic.

CASRA would like to thank all contributors to this initiative.

REFERENCES

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

[2] ICAO. (2021). Department for Transport. An effective security culture aviation as we recover from COVID-19. [PDF]

[3] ICAO. (2021). IATA. Improving performance through security culture. [PDF]

[4] Steiner-Koller, S. M., Bolfing, A., & Schwaninger, A. (2009). Assessment of X-ray image interpretation competency of aviation security screeners. In *43rd Annual 2009 International Carnahan Conference on Security Technology* (pp. 20-27). IEEE.. [PDF]

[5] Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers,* 31(1), (pp. 137-149).

[6] Hofer, F., & Schwaninger, A. (2004). Reliable and valid measures of threat detection performance in X-ray screening. *In 38th Annual 2004 International Carnahan Conference on Security Technology, 2004.* (pp. 303-308).

[7] Livingston, S. A. (2011). Item analysis. *In Handbook of test development (pp. 435-456).* Routledge.

[8] Wolfe, J. M., Brunelli, D. N., Rubinstein, J., & Horowitz, T. S. (2013). Prevalence effects in newly trained airport checkpoint screeners: Trained observers miss rare targets, too. *Journal of vision,* 13(3).

[9] Schwaninger, A., Koller, S.M, Bolfing, A. (2008). Assessment and Certification of X-Ray Image Interpretation Competency of Aviation Security Screeners. Report of VIA Project Work Package 6. [PDF]

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