COMPUTER-BASED TRAINING INCREASES EFFICIENCY IN X-RAY IMAGE INTERPRETATION BY AVIATION SECURITY SCREENERS

Stefan Michel, M Sc
Max Planck Institute
for Biological Cybernetics
Dept. Bülthoff
Spemannstrasse 38
D-72076 Tübingen
Germany

Jaap C. de Ruiter
TNO Defence, Security and Safety
Lange Kleiweg 137
PO Box 45
2280 AA Rijswijk
The Netherlands

Maarten Hogervorst, Dr.
TNO DSS Human Factors
Kampweg 5
3769 ZG Soesterberg
The Netherlands

Saskia M. Koller, M Sc
University of Zurich
Department of Psychology
Visual Cognition Research Group
Binzmuehlestrasse 14/22
8050 Zurich
Switzerland

Robert Moerland, M Sc
National Coordinator for Counterterrorism
Oranjebuitensingel
Postbus 16950
2500 BZ Den Haag
The Netherlands

Adrian Schwaninger, Dr.
University of Zurich
Department of Psychology
Visual Cognition Research Group
Binzmuehlestrasse 14/22
8050 Zurich
Switzerland

and

Max Planck Institute
for Biological Cybernetics
Dept. Bülthoff
Spemannstrasse 38
D-72076 Tübingen
Germany

Abstract – X-Ray screening of passenger bags is an essential component of airport security. Large investments into technology have been made in recent years. However, the most expensive equipment is of limited value, if the humans who operate it are not selected and trained appropriately. Scientific studies have shown that human performance in x-ray image interpretation depends critically on individual abilities and visual knowledge acquired through experience on the job and training. The aim of this study was to investigate the effect of adaptive computer-based training for increasing the detection of guns, knives, improvised explosive devices (IEDs), and other prohibited items. 97 airport security screeners of a European airport participated in this study. At the beginning of the project, all airport security screeners conducted the X-Ray Competency Assessment Test (X-Ray CAT). Thereupon they received adaptive computer-based training (CBT) for about 4 months. Then they conducted the X-Ray CAT the second time in the middle of the project. This was followed by about 4 months of CBT and a third test with X-Ray CAT at the end of the project. The goal was that each screener conducts at least one 20 minute training session per week. Substantial increases of detection performance were found as a result of training, which depended on the threat category (guns, IEDs, knives and other prohibited items). The largest training effects were found for IEDs. Additional analyses showed that training not only leads to an increase of detection performance but also results in faster response times when an x-ray image contains a threat object. Thus, recurrent CBT can be a powerful tool to increase efficiency in x-ray image interpretation by airport security screeners.

Index Terms — Airport security, human factors, computer-based training, x-ray screening.

I. INTRODUCTION

In response to the increasing threat of terrorist attacks, large investments were made into x-ray screening machines of the newest generation in order to inspect passenger baggage at airport security checkpoints. The last decision however is always made by a human operator (screener). The most expensive equipment is of little value if the screeners who operate it are not selected and trained appropriately. They have to be able to detect threat objects in passenger luggage within few seconds of inspection time.

Object shapes that are not similar to ones stored in visual memory are difficult to recognize ([1]; [2]; [3]). Detection of forbidden objects in x-ray images of passenger baggage depends on knowledge-based as well as on image based factors [4]. An airport security screener has to know which objects are prohibited and what they look like in an x-ray image. Some objects look quite different in x-ray images than in reality, for example an electric shock device. Other threat objects, like improvised explosive devices (IEDs), are rarely seen, in every day life as well as at the security checkpoint. Thus, it is not surprising that computer-based training (CBT) is very important to achieve and maintain a high detection performance, which is of special importance for detecting IEDs ([5]; [6]). Furthermore, threat objects can be depicted in an unfamiliar rotation in the baggage which can have a big impact on the detection performance. Based on research findings from object recognition ([2]) a large and representative image library of prohibited items depicted from different viewpoints is necessary to provide a good basis for training x-ray image interpretation competency. In addition to knowledge based factors, also image based factors play an important role. These can rather be attributed to the visual abilities of a person, that is, to the abilities to cope with image difficulty resulting from rotation of a threat object, superposition by other objects in the bag, and bag complexity ([4]; [7]; [8]; [9]; [10]).

A comparison of the detection performance of novice screeners with the one of trained aviation security screeners in an earlier study revealed a rather poor recognition of unfamiliar object shapes (e.g. self-defence gas spray, electric shock device etc.) in x-ray images for novices, whereas for trained aviation security personnel a much higher recognition performance was shown ([4]). [6] showed that adaptive CBT can be very effective to increase the detection of improvised explosive devices (IEDs) in x-ray images of passenger bags. [11] reported a better performance after training for the detection of knives in x-ray images.

II. METHOD AND PROCEDURE

The aim of this research project was to investigate to what extent recurrent adaptive CBT using X-Ray Tutor increases x-ray image interpretation competency. 97 airport security screeners of a European airport participated in this study. At the beginning of the project all airport security screeners conducted the X-Ray Competency Assessment Test (X-Ray CAT, [12]). Thereupon they received training for about 4 months, then conducted the X-Ray CAT a second time in the middle of the project. This was followed by about 4 months of training and a third test with X-Ray CAT at the end of the project. The goal was that each screener conducts at least one 20 minute training session per week.

A. X-Ray Competency Assessment Test (X-Ray CAT)

The X-Ray CAT is a standardized, reliable and valid instrument to measure x-ray image interpretation competency as defined by the principles and requirements specified in [13]. It contains 256 x-ray images of passenger baggage (see Figure 1).

Fig. 1: Example of an x-ray image of a passenger bag. The image on the right contains the prohibited item depicted separately on the bottom right.

Half of the bags contain a threat item, the other 128 bags are harmless. The threat items belong to the four categories guns, knives, IEDs and other prohibited items as defined in ECAC DOC 30. Each category is represented by 16 threat objects (8 visually similar pairs) and each object is depicted in the baggage in an easy and a difficult view. Of the visually similar pairs, only one item is used in training with X-Ray Tutor, while the other item is not used during training. A recent study showed that the performance improvements as a result of training with X-Ray Tutor generalize to visually similar objects not shown during training ([14]). While the easy view corresponds to the most usual (canonical) view, in the difficult view the threat object is rotated 85° either around the horizontal or the vertical axis. At test, airport security screeners have to decide for each bag whether it is OK (bag without threat item) or NOT OK (bag containing a threat item). Each image is depicted for a maximum of 15 seconds. Depending on how many images an airport security screener can visually inspect during 20 minutes, the test lasts about 2-3 sessions of 20 minutes. For more detailed information about the X-Ray CAT see [12] and [14].

B. X-Ray Tutor

X-Ray Tutor is a scientifically based training program. It is based on findings about how the human brain processes visual information in order to recognise objects in different views, when superimposed by other objects, and depending on bag complexity ([3]). The training is individually adaptive, that is, it automatically adapts to the performance of individual airport security screeners. X-Ray Tutor automatically combines images of fictional threat items with x-ray images of passenger bags. This is performed by an individually adaptive algorithm, which takes into account the rotation of threat objects, the superposition by other objects in the bag, and bag complexity resulting from clutter and transparency of the objects in the baggage. X-Ray Tutor 2.0 contains a large image library of threat objects that are depicted in different standardized views. Most of the objects can be depicted from up to 72 different viewpoints, which allows training screeners to detect threat objects independent of rotation. This image library was built in close collaboration with experts of Zurich State Police, Airport Division, and it is being extended continuously.

During training with X-Ray Tutor, x-ray images of bags are depicted on the screen for 15 seconds (standard setting). Screeners have to decide whether the bag is OK (i.e. it contains no threat object) or whether it is NOT OK (i.e. it contains a threat object). After each response, a feedback is provided informing the screener whether his/her response was correct. If the bag contained a threat object the user can view detailed information and a real image of the threat object. For further information on X-Ray Tutor see [3].

III. RESULTS AND DISCUSSION

In this study, the performance of screeners to detect threat objects in the X-Ray CAT as well as the time needed for detecting the threat objects (i.e. reaction time) has been analyzed. An effect of training could imply an increase in detection performance and/or a decrease in reaction time. Furthermore, the effect of object viewpoint has been analyzed, i.e. a possible difference in detection performance of threat objects depending on the rotation with which they are depicted in the image.

A. Detection Performance

Detection performance in the X-Ray CAT was analysed using d', a widely used measure of sensitivity based on signal detection theory ([15]; [16]; [17]). The d' measure takes into account the hit rate as well as the false alarm rate. It can be calculated by the following formula: d' = z(H) – z(FA), whereas H is the hit rate, FA the false alarm rate and z refers to the z-transform.

The hit rate indicates how often a person correctly judges a bag as being NOT OK proportionately to all bags containing a threat object. The false alarm rate indicates how often a person wrongly judges a bag as being NOT OK proportionately to all bags containing no threat object. In this study, actual performance values are not reported due to security reasons. However, effect sizes are reported for all relevant analyses and interpreted based on [18].

1) Effect of Training: Figure 2 shows the detection performance and the standard deviation[1] for the easy view of the threat objects in each category for all three test measurements. Guns were detected best in all three tests and objects of the categories IEDs and "Other" were detected worst. Substantial increases of detection performance were found, which depended on the threat category. The largest training effects were found for IEDs. Good performance was achieved for IEDs after the two training blocks of 4 months each. The aim was that all screeners conduct at least one 20 minute training session per week, which was achieved on average. Note that there are large differences between screeners as can be seen by the large standard deviations (thin lines in Figure 2). Some screeners achieved very good performance for all types of threat objects after the two training phases. This is mainly due to differences in the amount of training. While some screeners did only a few trainings over several months, other screeners did several training sessions per week and achieved very high performance increases.

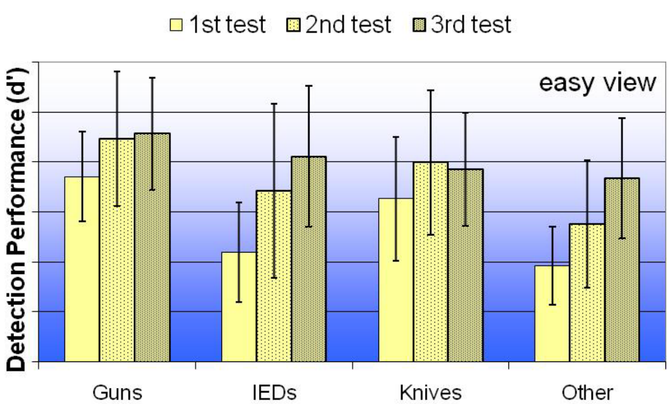

Fig. 2: Detection performance d' and standard deviations for easy views broken up by threat category and test date. *Note*: For security reasons d' scores are not indicated in the figure.

An analysis of variance (ANOVA) for repeated measures using d' scores for easy view with the within-participant factors test date (first, second, third) and category (guns, knives, IEDs, other) revealed large main effects of test date $\eta^2 = .36$, $F(2, 192) = 53.82$, $p < .001$, and category $\eta^2 = .57$, $F(3, 288) = 128.54$, $p < .001$, and a large two-way interaction of test date and category $\eta^2 = .15$, $F(6, 576) = 16.80$, $p < .001$. These results confirm that CBT with X-Ray Tutor result in large performance increases of x-ray screeners, especially regarding the detection of IEDs and other threat items.

2) Effect of Object Viewpoint: Figure 3 shows the results for detection performance d' and standard deviations when threat objects were depicted from a difficult viewpoint. Figure 4 depicts the comparison between the detection performance for objects in easy view and difficult view. It shows that detection of threat objects is much easier in frontal (or canonical view) than when depicted from a difficult viewpoint. Large training effects have been found for difficult views (cf. knives) as well as for objects that are rarely seen in every day life (IEDs and other threat items). The reason for this large performance increase is the fact that X-Ray Tutor contains a large threat image library in which objects are depicted in many different viewpoints [3]. X-Ray Tutor trains each screener individually to become able to detect all types of threat objects even if they are shown from a difficult viewpoint. This is the reason why such large training effects were found for objects shown in difficult view.

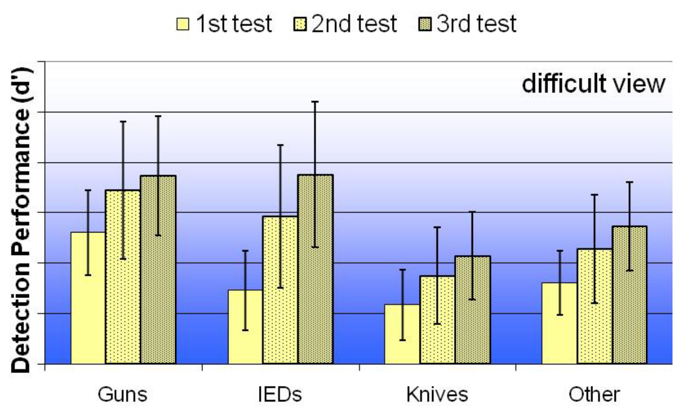

Fig. 3: Detection performance d' and standard deviations for difficult views broken up by threat category and test date. *Note*: For security reasons d' scores are not indicated in the figure.

An analysis of variance (ANOVA) for repeated measures using d' scores for difficult view with the within-participant factors test date (first, second, third) and category (guns, knives, IEDs, other) revealed a large main effect of test date $\eta^2 = .46$, $F(2, 192) = 82.22$, $p < .001$, a large main effect of category $\eta^2 = 68$, $F(3, 288) = 200.32$, $p < .001$, and a large two-way interaction of test date and category $\eta^2 = .15$, $F(6, 576) = 16.94$, $p < .001$. These results provide further evidence for the effectiveness of recurrent CBT with X-Ray Tutor, which results in large performance increases for detecting threat items in x-ray images.

An additional analysis of variance (ANOVA) for repeated measures using d' scores for both views (easy and difficult) with the within-participant factors test date (first, second, third), view

[1] The standard deviation represents the range of dispersion around the mean of the data and indicates the range of individual differences between the tested airport security screeners.

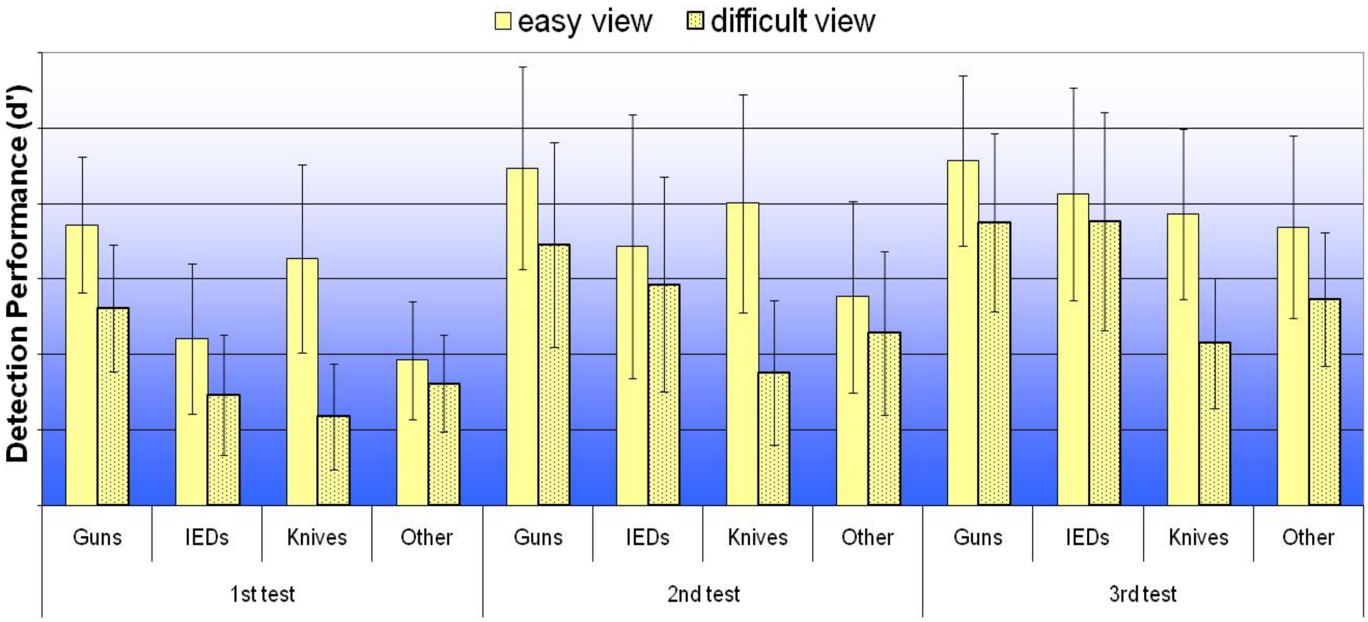

Fig. 4: Detection performance d' and standard deviations broken up by threat category, view (easy vs. difficult), and test date. *Note*: For security reasons d' scores are not indicated in the figure.

(easy vs. difficult), and category (guns, knives, IEDs, other) was conducted to examine the effect of viewpoint on the detection performance of screeners. There was a large main effect of test date $\eta^2 = .43$, $F(2, 192) = 73.23$, $p < .001$, a large main effect of view (easy vs. difficult) $\eta^2 = .89$, $F(1, 96) = 804.15$, $p < .001$, and a large main effect of category $\eta^2 = .66$, $F(3, 288) = 182.39$, $p < .001$. The following interactions were significant as well with large effects for the two-way interactions between test date and category $\eta^2 = .19$, $F(6, 576) = 22.34$, $p < .001$, and between viewpoint and category $\eta^2 = .56$, $F(3, 288) = 120.00$, $p < .001$. There was also a medium effect for the three-way interaction of test date, view and category, $\eta^2 = .08$, $F(6, 576) = 7.87$, $p < .001$.

These results show that recurrent CBT with X-Ray Tutor is very effective to train screeners to detect threat objects even if they are depicted from an unusual viewpoint.

B. Reaction Times

For each response, reaction time (RT) was measured, i.e. the time between x-ray image onset and the time a response was provided by the screener (OK or NOT OK button). Figure 5 shows the RTs of all hits (correctly judged images as NOT OK) of all three test dates broken up by threat category. RTs decreased as a result of training, especially from the first to the second test. However, there were also large differences between individual airport security screeners (cf. large standard deviations), possibly due to differences in amount of training. Other RT data (misses, correct rejections and false alarms) are not discussed here but can be provided upon request.

An analysis of variance (ANOVA) for repeated measures using RTs with the within-participant factors test date (first, second, third) and category (guns, knives, IEDs, other) revealed a large main effect of test date $\eta^2 = .47$, $F(2, 192) = 85.01$, $p < .001$, a large main effect of category $\eta^2 = .44$, $F(3, 288) = 76.42$, $p < .001$, and a large two-way interaction of test date and category $\eta^2 = .16$, $F(6, 576) = 18.79$, $p < .001$.

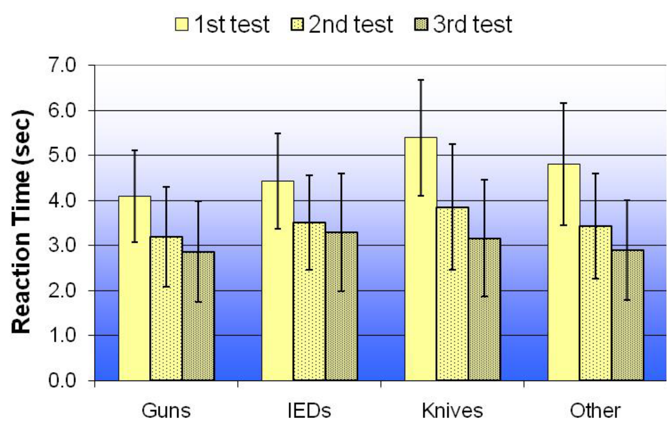

Fig. 5: Reaction times and standard deviations of hits (correctly answered as NOT OK) broken up by threat category and test date.

IV. SUMMARY AND CONCLUSIONS

This study has shown substantial increases of airport security screeners' x-ray image interpretation competency as a result of recurrent adaptive CBT using X-Ray Tutor. The largest increase was found for the detection of IEDs. After two training phases of 4 months each, detection was almost as good as detection of guns. There were also large effects of viewpoint. Objects shown from a difficult rotation are more difficult to recognize (especially knives) than when depicted from a non-rotated canonical view. However, the effect of viewpoint can be compensated by training. At the third test, the difficult views were recognized much better than before training started (first test). There were also large differences between screeners. While some screeners did only a few trainings per month, others did several training sessions per week and achieved very large performance increases. A large effect of training was also found in the reaction times. Screeners could reduce the time needed to detect a threat object significantly. More detailed analyses showed that the increase in the detection performance was mainly due to an increase in the hit rate (as opposed to a decrease in the false alarm rate), which means, that a speed-accuracy trade off can be ruled out here.

It is not surprising that before training, d' scores for IEDs and also for other threat categories were substantially smaller than for guns. The IEDs used in this study are quite sophisticated threat objects using components that are often not known to screeners without enhanced training in IED detection. For other threat items probably this knowledge based factor comes into play as well. Screeners first have to learn which objects are prohibited and what they look like in x-ray images. If an effective CBT is used for recurrent training, a large increase in detection performance can be achieved (see also [3]). This shows that the detection of IEDs and other threat items it not difficult per se, but rather depending on the training of screeners.

A recent study by [14] showed that training not only has an effect on the detection of trained object views – which then are available in visual memory – but also generalizes to similar looking but untrained views of other objects. This transfer effect was revealed for all threat categories, i.e. for guns, IEDs and other threat items. For knives no transfer effect was found. This could be due to the shape of the knives. On one hand, knives show less diagnostic features which play an important role in object recognition compared to objects from other categories. These few diagnostic features might also get lost when a knife is rotated. On the other hand, the visual similarity of knives to harmless everyday objects (e.g., pen) is substantial. These factors could impede detectability and trainability of knives and ultimately might have resulted in rather small training and transfer effects for knives. They might also be an explanation for the fact that for knives there is the largest viewpoint effect compared to objects from other categories (see Figure 4). While objects of other threat categories in a rotated view usually still show many diagnostic features and also, due to their larger surface, more information in general, for knives much information might be lost with a high rotation angle. Therefore, the detection and the discrimination of knives and harmless objects is hindered. However, rather than being a category generalisation of the gained knowledge, the transfer effect could also have resulted because of a large similarity of the object pairs. This would imply a specific-token familiarity to be the reason for the transfer as [19] suggest. If a specific-token familiarity would apply to the recognition of objects then learned knowledge about an object could not be transferred to an object of the same category but only to objects with the same specific tokens. For a more detailed discussion of this issue see [14] and [19].

Overall, these results are fully consistent with earlier results ([3]; [5]; [6]; [20]) and show that adaptive CBT such as X-Ray Tutor can be a powerful tool to increase efficiency in x-ray image interpretation by airport security screeners.

V. ACKNOWLEDGEMENTS

This research was financially supported by the European Commission Leonardo da Vinci Programme (VIA Project, DE/06/C/F/TH-80403) and the Ministry of Justice, The Netherlands.

VI. REFERENCES

[1] M. Graf, A. Schwaninger, C. Wallraven, and H.H. Bülthoff, "Psychophysical results from experiments on recognition & categorisation," *Information Society Technologies (IST) programme, Cognitive Vision Systems – CogVis (IST-2000-29375),* 2002.

[2] A. Schwaninger, "Objekterkennung und Signaldetektion," In: B. Kersten & M.T. Groner (Eds.), *Praxisfelder der Wahrnehmungspsychologie* Bern: Huber, 2004, pp 106-130.

[3] A. Schwaninger, "Computer based training: a powerful tool to the enhancement of human factors," *Aviation Security International*, vol FEB/2004, pp 31-36.

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

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

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

[7] D. Hardmeier, F. Hofer, and A. Schwaninger, "The x-ray object recognition test (x-ray ort) – a reliable and valid instrument for measuring visual abilities needed in x-ray screening," in *IEEE ICCST Proceedings,* vol 39, 2005, pp 189-192.

[8] D. Hardmeier, F. Hofer, and A. Schwaninger, "Increased detection performance in airport security screening using the x-ray ort as pre-employment assessment tool," in *Proceedings of the 2nd International Conference on Research in Air Transportation, ICRAT 2006, Belgrade, Serbia and Montenegro,* June 24-28, 2006, pp 393-397.

[9] F. Hofer, D. Hardmeier, and A. Schwaninger, "Increasing airport security using the X-ray ORT as effective pre-employment assessment tool," in *Proceedings of the 4th International Aviation Security Technology Symposium, Washington, D.C., USA,* November 27 – December 1, 2006, pp 303-308.

[10] A. Schwaninger, “Evaluation and selection of airport security screeners,” *AIRPORT*, vol 02/2003, pp 14-15.

[11] J.S. McCarley, A.F. Kramer, C.D. Wickens, E.D. Vidoni, and W.R. Boot, “Visual skills in airport screening, *Psychological Science*, vol 15(5), pp 302-306, 2004.

[12] S. Koller and A. Schwaninger, “Assessing X-ray image interpretation competency using the X-Ray Tutor Competency Assessment Test X-Ray CAT,” in *Proceedings of the 2nd International Conference on Research in Air Transportation, ICRAT 2006, Belgrade, Serbia and Montenegro,* June 24-28, 2006, pp 399-402.

[13] A. Schwaninger, A.P. Bridges, C. Drury, F. Durinckx, P. Durrant, T. Hodge, F. Hofer, R. Jongejan, R.L. Maguire, A. McClumpha, E. Neiderman, C. Steinmann, and W. Wüest, *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), 2005.

[14] S.M. Koller, D. Hardmeier, S. Michel, and A. Schwaninger, “Investigating training and transfer effects resulting from recurrent CBT of x-ray image interpretation,” in D.S. McNamara and J.G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society,* 2007, pp 1181-1186.

[15] D.M. Green, and J.A. Swets, *Signal Detection Theory and Psychophysics*, New York: Wiley, 1966.

[16] F. Hofer, and A. Schwaninger, “Reliable and valid measures of threat detection performance in X-ray screening,” in *IEEE ICCST Proceedings*, vol 38, pp 303-308, 2004.

[17] N.A. MacMillan, and C.D. Creelman, *Detection theory: A user’s guide*, Cambridge: University Press, 1991.

[18] J. Cohen, *Statistical power analysis for the behavioural sciences*, New York: Erlbaum, Hillsdale, 1988.

[19] J.D. Smith, J.S. Redford, L.C. Gent, and D.A. Washburn, “Visual search and the collapse of categorization,” *Journal of Experimental Psychology: General*, vol 134(4), pp 443-460, 2005.

[20] A. Schwaninger, “Training of airport security screeners,” *AIRPORT*, vol 05/2003, pp 11-13.