

Eine erste explorative Studie zur Relevanz von Wissen über das Aussehen von Alltagsgegenständen bei der Röntgenbildbeurteilung in der Luftsicherheit

Nicole HÄTTENSCHWILER, Stefan MICHEL, Milena KUHN,
Sandrina RITZMANN, Adrian SCHWANINGER

*Institut Mensch in komplexen Systemen (MikS), Hochschule für
Angewandte Psychologie, Fachhochschule Nordwestschweiz (FHNW),
Riggenbachstrasse 16 4600 Olten, Schweiz*

Kurzfassung: Es wird angenommen, dass das Wissen über das Aussehen von Alltagsgegenständen Luftsicherheitsassistenten helfen könnte, effizienter zu arbeiten, indem weniger Fehlalarme generiert werden. Im ersten Experiment wurde der Zusammenhang zwischen dem Wissen über Alltagsgegenstände und der Fehlalarmrate mit 15 Screenern untersucht, wobei sich ein sign. negativer Zusammenhang zeigte. In einem zweiten Experiment wurde die Effektivität eines E-Learning Moduls zur Aneignung von Wissen über Alltagsgegenständen mit 30 Novizen evaluiert. Die Auswertungen zeigten, dass ein E-Learning Modul eine effektive und effiziente Methode zur Aneignung von Wissen über Alltagsgegenstände darstellt. Die Relevanz von Wissen über Alltagsgegenstände sowie entsprechendes Training werden diskutiert.

Schlüsselwörter: Alltagsgegenstände, Luftsicherheit, Mensch-Maschinen-Interaktion, Röntgenbildanalyse, E-learning

1. Einleitung

Die Luftsicherheit hat im letzten Jahrzehnt stark an Bedeutung gewonnen, und es wurden grosse Investitionen in die Optimierung der Sicherheitskontrollen investiert. Luftsicherheitsassistenten (Screener) analysieren Gepäckstücke anhand von Röntgenbildern um sicherzustellen, dass keine verbotenen oder gefährlichen Gegenstände an Bord eines Flugzeuges gelangen. Mehrere Studien konnten zeigen, dass computer-basiertes Training (CBT) eine effiziente und effektive Methode darstellt, um die Erkennung von verbotenen oder gefährlichen Gegenständen in Röntgenbildern zu steigern (z.B. Schwaninger & Hofer, 2004; Michel et al., 2007; Schwaninger et al. 2007). Die Resultate dieser Studien sprechen dafür, dass das Erlernen des Aussehens von verbotenen oder gefährlichen Gegenständen in Röntgenbildern mittels CBT eine wichtige Voraussetzung für effektives und effizientes Arbeiten bei der Röntgenbildanalyse an Luftsicherheitskontrollstellen darstellt. Bis heute gibt es gemäss unserem Kenntnisstand jedoch noch keine Studie, die untersucht hat, ob das (visuelle) Wissen über harmlose Alltagsgegenstände für eine effektive und effiziente Röntgenbildanalyse ebenfalls relevant ist. Es wird angenommen, dass fehlendes Wissen über Alltagsgegenstände dazu führen könnte, dass ein Alltagsgegenstand mit einem verbotenen Gegenstand verwechselt wird. Dies würde dazu führen, dass ein Screener ein Gepäckstück fälschlicherweise als gefährlich einstuft und somit einen Fehlalarm generiert. Fehlalarme resultieren in einer längeren Wartezeit für die Passagiere, weil Gepäckstücke weiter untersucht

werden müssen durch manuelle Nachkontrolle und den Einsatz eines Sprengstoffdetektionssystems (explosive trace detection system, ETD). Daher kann davon ausgegangen werden, dass das Wissen über Alltagsgegenstände eine hohe operative Relevanz aufweist, weil dadurch die Effizienz an der Luftsicherheitskontrollstelle gesteigert werden könnte (Sterchi & Schwaninger, 2015).

Diese erste explorative Studie untersuchte anhand von zwei Experimenten die Relevanz des Wissens über harmlose Alltagsgegenstände bei der Röntgenbildbeurteilung in der Luftsicherheit. Dafür wurde ein Alltagsgegenstandstest (AGT) entwickelt, um dieses Wissen bei Screenern zu ermitteln. In Experiment 1 wurden bivariate und partielle Korrelationen gerechnet, um den Zusammenhang zwischen dem Wissen über Alltagsgegenstände und der Fehlalarmrate in einem simulierten Röntgenbildtest (SRBT) zu untersuchen. Experiment 2 untersuchte, inwiefern ein E-Learning Modul als effektive und effiziente Methode genutzt werden kann, um Wissen über Alltagsgegenstände zu vermitteln.

2. Experiment 1

Das Ziel von Experiment 1 war die Untersuchung des Zusammenhangs zwischen dem Wissen über Alltagsgegenstände und der Fehlalarmrate bei der Röntgenbildanalyse. Es wurden ein Alltagsgegenstandstest (AGT) und ein simulierter Röntgenbildtest (SRBT) mit Screenern durchgeführt.

2.1 Methoden und Prozedere

15 erfahrene Screener eines grossen internationalen Flughafens nahmen an Experiment 1 teil. Das durchschnittliche Alter der Screener beträgt 39.8 Jahre (SD = 10.13) und die durchschnittliche Arbeitserfahrung 6.03 Jahre (SD = 4.44).

Der Alltagsgegenstandstest (AGT) umfasste 32 Röntgenbilder von Handgepäckstücken, wobei auf jedem Bild drei Gegenstände durch einen roten Rahmen markiert waren. Bei den markierten Gegenständen handelt es sich entweder um einen verbotenen Gegenstand oder einen harmlosen Alltagsgegenstand. Insgesamt umfasste der Test 17 verbotene Gegenstände und 79 harmlose Alltagsgegenstände. Die Screener hatten die Aufgabe, diese Gegenstände zuerst zu kategorisieren und anschliessend zu benennen. Die zu kategorisierenden Gegenstände konnten den Kategorien "Alltagsgegenstand", "Unkonventionelle Spreng- und Brandvorrichtung (USBV)" oder "anderer verbotener Gegenstand"

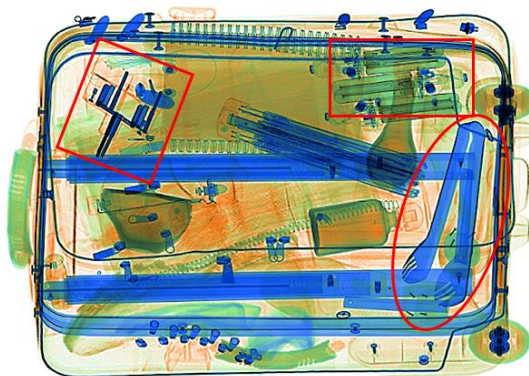


Abbildung 1: Darstellung eines Röntgenbildes eines Gepäckstücks aus dem Alltagsgegenstandstest (AGT) mit rot markierten Objekten.

zugeordnet werden. Anschliessend musste der Gegenstand möglichst treffend benannt werden. Die Bearbeitung des Tests dauerte 45 bis 60 Minuten.

Beim simulierten Röntgenbildtest (SRBT) mussten die Screener bei 160 Röntgenbildern entscheiden, ob diese einen verbotenen Gegenstand enthielten (Nicht OK) oder harmlos (OK) waren. Die Hälfte aller Röntgenbilder enthielt einen verbotenen Gegenstand aus den Kategorien Messer, Schusswaffen, USBV oder andere verbotene Gegenstände. Die verbotenen Gegenstände wurden einmal in einer einfachen und einmal in einer rotierten Ansicht gezeigt. Die Bearbeitung des Tests dauerte ca. 30 Minuten.

2.2 Resultate und Diskussion

Im AGT konnten 96.19% (SD = 1.67) der Gegenstände korrekt kategorisiert und Alltagsgegenstände zu 82.07% (SD = 11.19) korrekt benannt werden. Als Fehlalarm im SRBT zählte, wenn ein Bild fälschlicherweise als gefährlich (Nicht OK) eingestuft wurde, obwohl dieses keinen verbotenen Gegenstand enthielt. Sie wird aus der Anzahl Fehlalarme, geteilt durch die Anzahl aller Röntgenbilder, die keinen verbotenen Gegenstand enthielten, berechnet. Die durchschnittliche Fehlalarmrate der Screener lag bei $M = .037$ (SD = .031). Eine einseitige Pearson-Korrelation zeigte einen signifikanten negativen Zusammenhang $r(13) = -.471$, $p = .038$ zwischen "% richtige Benennung im AGT" und der "Fehlalarmrate im SRBT" (siehe Abb. 2).

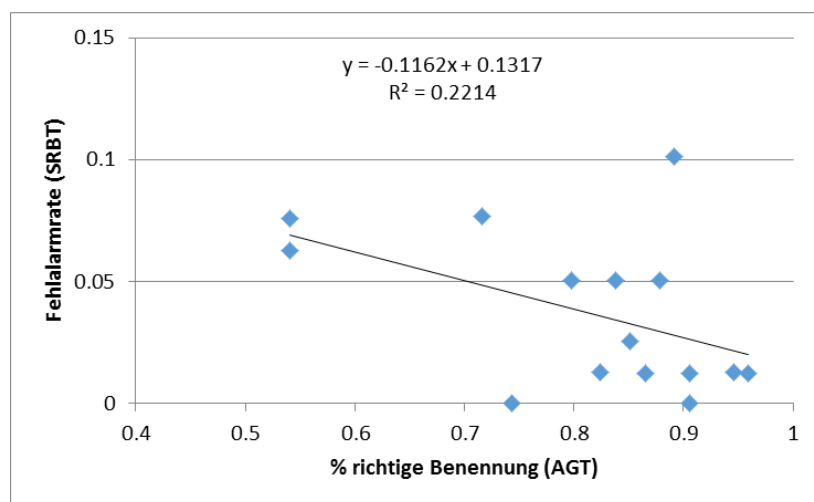


Abbildung 2: Zusammenhang zwischen Fehlalarmrate im SRBT und Prozentsatz richtiger Benennungen im AGT.

Die Signaldetektionstheorie (für detailliertere Informationen siehe Green & Swets, 1966) erklärt, dass eine Variation in der Fehlalarmrate vom Antwortverhalten abhängen kann. Das Antwortverhalten widerspiegelt dabei die Tendenz eines Screeners, mit welcher Häufigkeit angegeben wird, dass sich ein verbotener Gegenstand (ein Signal) im Gepäckstück befindet. Dieses kann mit der Variable "Kriterium" $c = -0.5[z(H) + z(FA)]$ gemessen werden, wobei $z(H)$ die z-transformierte Trefferrate und $z(FA)$ die z-transformierte Fehlalarmrate darstellt. Die Untersuchung des Zusammenhangs zwischen der Fehlalarmrate und dem Kriterium im SRBT ergab eine hohe signifikante negative Korrelation von $r(13) = -.870$, $p < .001$.

Um den Zusammenhang zwischen dem Wissen über Alltagsgegenstände und der Fehlalarmrate ohne den Einfluss des Kriteriums zu bestimmen, wurde eine partielle Korrelation berechnet mit dem Kriterium als Kontrollvariable. Es zeigte sich ein

signifikanter Zusammenhang zwischen dem Wissen über Alltagsgegenstände und der Fehlalarmrate, $r(13) = -.492$, $p = .037$; die partielle Korrelation war damit sogar geringfügig höher als die weiter oben erwähnte einseitige Pearson-Korrelation ($r(13) = -.471$, $p = .038$).

Diese Ergebnisse sind konsistent mit der Annahme, dass ein fundiertes Wissen über Alltagsgegenstände in Röntgenbildern zu einer tieferen Fehlalarmrate bei der Röntgenbildanalyse führt, was sich positiv auf die Effizienz an der Luftsicherheitskontrollstelle auswirken dürfte, weil dadurch weniger Gepäckstücke nachkontrolliert werden müssen.

3. Experiment 2

Die Ergebnisse aus Experiment 1 sprechen dafür, dass ein fundiertes Wissen über Alltagsgegenstände für eine effiziente Röntgenbildanalyse relevant ist. Zudem konnte gezeigt werden, dass Screener über ein relativ gutes Wissen über Alltagsgegenstände verfügen, jedoch grosse interindividuelle Unterschiede bestehen. In Experiment 2 wurde nun untersucht, wie gut Novizen harmlose Alltagsgegenstände benennen können, und ob dieses Wissen effizient und effektiv trainiert werden kann.

3.1 Methoden und Prozedere

An Experiment 2 nahmen 30 Novizen teil (50% weiblich) mit einem durchschnittlichen Alter von 31.1 Jahren ($SD = 11.87$) und keiner Arbeitserfahrung im Bereich der Röntgenbildanalyse. Die Novizen wurde zuerst in Bezug auf ihre visuell-kognitiven Fähigkeiten mittels eines für die Personalauswahl von Screenern entwickelten Röntgenbildtests (X-Ray Object Recognition Test: ORT) getestet (für weitere Informationen zum X-Ray Object Recognition Test, siehe Hardmeier et al., 2006). Basierend auf den Daten des ORTs wurden im Anschluss alle Versuchsteilnehmenden in zwei gleiche Gruppen eingeteilt, eine Experimental- und Kontrollgruppe. Mittels eines Test-Retest-Designs lösten beide Gruppen einen adaptierten Alltagsgegenstandstest (aAGT) aus Experiment 1. Die Experimentalgruppe erhielt im Anschluss ein Training zum Erlernen von Alltagsgegenständen in Röntgenbildern (X-Ray Introduction - XRI) und löste anschliessend den aAGT erneut. Die Kontrollgruppe löste den aAGT zweimal wobei sie kein Training zum Erlernen von Alltagsgegenständen in Röntgenbildern absolvierten. Der zeitliche Abstand zwischen der ersten (Prä) und zweiten (Post) Durchführung des aAGT betrug bei beiden Gruppen eine Woche. Beim aAGT handelt es sich um die gleiche Version wie in Experiment 1, wobei nur noch Alltagsgegenstände mit einem roten Rahmen markiert waren und die rote Markierung der verbotenen Gegenstände entfernt wurden. Die Novizen mussten die Alltagsgegenstände lediglich benennen und nicht kategorisieren. Das Training (XRI) fand in Form eines E-Learning Moduls statt. Dabei wurden in mehreren Lektionen Alltagsgegenstände in drei aufeinanderfolgenden Schritten gelernt: In einem ersten Schritt wurden Alltagsgegenstände als Foto und als Röntgenbild gezeigt und vergleichend dargestellt. In einem zweiten Schritt wurde, um einen realistischen Kontext zu schaffen, der Alltagsgegenstand in einem Gepäckstück als Röntgenbild gezeigt. Zuletzt wurde dem Anwender die Möglichkeit geboten, sein neu erlerntes Wissen zu testen.

3.2 Resultate und Diskussion

Die Resultate vom ersten (Prä) und zweiten Zeitpunkt (Post) der Durchführung des aAGT können der Abbildung 3 entnommen werden. Bei der Prä-Testung unterschieden sich die Gruppen bzgl. der prozentualen Anzahl korrekt benannter Alltagsgegenstände nicht signifikant $t(28) = -.377, p = .709$. Bemerkenswert ist auch, dass die Hälfte der Gegenstände bereits ohne Training von den Novizen richtig benannt werden konnten. Beim zweiten Testzeitpunkt (Post) verbesserte sich die Leistung der Experimentalgruppe im Vergleich zum ersten Testzeitpunkt (Prä) signifikant $t(28) = -2.549, p = .008$, wobei die Leistung der Kontrollgruppe sich nicht signifikant verändert hatte $t(28) = -1.305, p = .101$. Die Berechnungen der Haupteffekte und der Interaktion wurden mittels einer univariaten 2x2 ANOVA gerechnet mit der abhängigen Variable "Prozentsatz korrekter Benennung", dem Inter-Subjekt-Faktor Testzeitpunkt (Prä vs. Post) und dem Zwischen-Subjekt Faktor Gruppe (Experimental- vs. Kontrollgruppe). Es wurde kein signifikanter Haupteffekt der Gruppe gefunden $F(1, 28) = 1.26, p = .272$, jedoch ein signifikanter Effekt des Faktors Testzeitpunkt $F(1, 28) = 77.89, p < .001$, sowie eine signifikante Interaktion der Variablen Testzeitpunkt x Gruppe $F(1, 28) = 9.42, p = .005$.

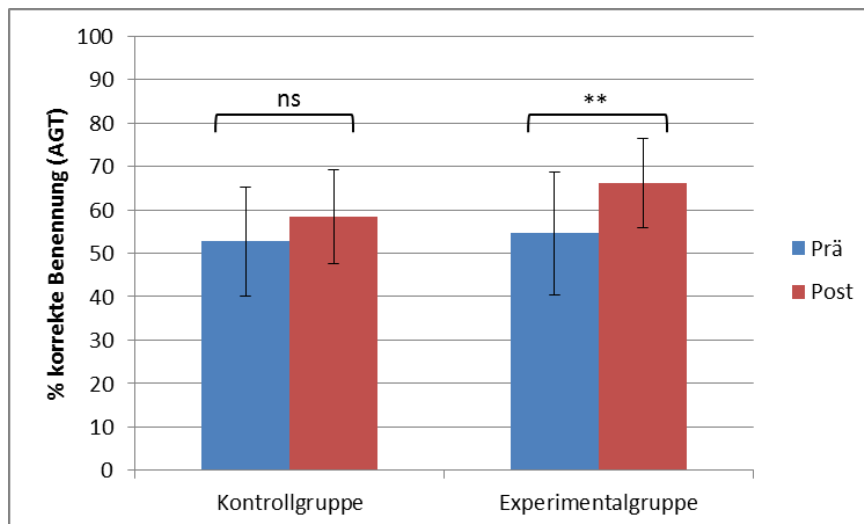


Abbildung 3: Mittelwerte und Standardabweichungen der % richtig benannter Alltagsgegenstände aus dem ATG für die Kontroll- und die Experimentalgruppe, vor und nach dem Absolvieren des Trainings mittels XRI.

Diese Resultate zeigen, dass bereits ein relativ kurzes E-Learning Modul das Wissen über Alltagsgegenstände signifikant verbessern und somit als effektives und effizientes Training eingesetzt werden kann.

4. Schlussfolgerung

Für eine effiziente Röntgenbildanalyse an der Luftsicherheitskontrollstelle ist eine möglichst tiefe Fehlalarmrate wünschenswert, da dadurch weniger Gepäckstücke einer zusätzlichen Kontrolle unterzogen werden müssen. Ob dabei das Wissen über Alltagsgegenstände eine Rolle spielt, wurde bisher noch nicht untersucht. Daher wurden in dieser Studie zwei Experimente durchgeführt, um erste Erkenntnisse zur Relevanz über das Wissen von Alltagsgegenständen bei der Röntgenbildanalyse in

der Luftsicherheit zu erhalten. Der angenommene negative Zusammenhang zwischen dem Prozentsatz korrekter Benennungen von Alltagsgegenständen und der Fehlalarmrate in einem simulierten Röntgenbildtest wurde mittels Korrelationsanalysen bestätigt. Dies ist konsistent mit der Annahme, dass ein fundiertes Wissen über Alltagsgegenstände in Röntgenbildern dazu führt, dass weniger Fehlalarme bei der Röntgenbildbeurteilung entstehen und dadurch die Effizienz bei der Röntgenbildanalyse zunimmt. Einschränkend ist darauf hinzuweisen, dass Korrelationsanalysen keinen Schluss auf Kausalitäten zulassen, selbst wenn partielle Korrelationen (im vorliegenden Fall mit Kriterium c als Kontrollvariable) berechnet werden. Der postulierte kausale Zusammenhang zwischen Wissen über Alltagsgegenstände in Röntgenbildern und Effizienz bei der Röntgenbildanalyse (gemessen mit der Fehlalarmrate) müsste in zukünftigen Studien mit einem experimentellen Design und einer grösseren Stichprobe weiter untersucht werden.

Die Resultate aus Experiment 2, bei dem Novizen ein Training über das Wissen von Alltagsgegenständen absolvierten, sind vielversprechend. Nach Absolvieren dieses Trainings zeigte die Experimentalgruppe eine signifikant höhere Leistung im Vergleich zur Kontrollgruppe. Vor allem für neue Screener könnten diese Ergebnisse vielversprechend sein, damit diese schnellstmöglich effizient an der Luftsicherheitskontrollstelle arbeiten können. Obwohl erfahrenen Screener bereits gute Resultate beim Benennen von Alltagsgegenständen erzielten, gab es grosse interindividuelle Unterschiede. Aus diesem Grund sollten Kosten-Nutzen-Analysen in Betracht gezogen werden, um abzuschätzen, ob das Training von Alltagsgegenständen nicht nur für die Erstausbildung sondern auch im Rahmen der gesetzlich vorgeschriebenen kontinuierlichen Weiterbildung angeboten werden soll.

Zusammenfassend kann festgehalten werden, dass die Ergebnisse dieser Studie die Annahme stützen, dass Wissen über Alltagsgegenstände zusammen mit dem Wissen über verbotene Gegenstände einen positiven Effekt auf die Effektivität und Effizienz bei der Röntgenbildanalyse an Luftsicherheitskontrollstellen hat.

5. Literatur

- Green DM, Swets JA (1966) Signal Detection Theory and Psychophysics. New York: Wiley.
- Hardmeier D, Hofer F, Schwaninger A (2005) The x-ray object recognition test (x-ray ort) – a reliable and valid instrument for measuring visual abilities needed in x-ray screening. IEEE ICCST Proceedings, 39: 189-192.
- Michel S, de Ruiter J, Hogervorst M, Koller S, Moerland R, Schwaninger A (2007) Computer-based training increases efficiency in x-ray image interpretation by aviation security screeners. Proceedings of the 41st Carnahan Conference on Security Technology, Ottawa, October 8-11, 2007.
- Schwaninger A, Hofer F (2004) Evaluation of CBT for increasing threat detection performance in X-ray screening. In: Morgan K, Spector MJ (Eds.) The Internet Society 2004, Advances in Learning, Commerce and Security, Wessex: WIT Press: 147-156.
- Schwaninger A, Hofer F, Wetter O (2007) Adaptive computer-based training increases on the job performance of x-ray screeners. Proceedings of the 41st Carnahan Conference on Security Technology, Ottawa, October 8-11, 2007.
- Sterchi Y, Schwaninger A (2015) Optimizing the introduction of EDS in cabin baggage screening: A first simulation of the effect on throughput. Proceedings of the 46th IEEE International Carnahan Conference on Security Technology, Taipei, September 21-24, 2015.