Ergonomics
Vol. 52, No. 6, June 2009, 644–656

Change of search time and non-search time in X-ray baggage screening due to training

Saskia M. Koller[a,b], Colin G. Drury[c] and Adrian Schwaninger[b,d]*

[a]*Department of Psychology, University of Zurich, Binzmühlestrasse 14/22, 8050 Zurich, Switzerland;* [b]*School of Applied Psychology, University of Applied Sciences Northwestern Switzerland (FHNW), Riggenbachstrasse 16, 4600 Olten, Switzerland;* [c]*Department of Industrial and Systems Engineering, University at Buffalo: SUNY, 438 Bell Hall, Buffalo, NY 14260 USA;* [d]*Department of Psychology, University of Zurich, Binzmühlestrasse 14/22, 8050 Zurich, Switzerland*

As found in studies of aircraft structural inspection, the time used for judging if a part of an aircraft shows tiny cracks is composed of search time, used for actively scanning, and non-search time, used for matching and decision while fixating a region of interest (Drury *et al.* 1997). These findings can be applied to detection of threats by X-ray screening of passenger bags at airports. To investigate whether search time and non-search time change when an experienced screener is given additional training in recognising threat objects in passenger bags, data from a European airport were analysed. A comparison of detection performance and reaction time between two large groups of screeners, one trained for 6 months, shows a large impact of training on overall performance and on both search and non-search components of the task. There was also a small but consistent decline in performance measures with screener age. This study shows a way to localise the effect of training on threat detection performance for aviation security screening. Analysis of the time needed for screening each passenger bag showed that training had a significant effect, particularly on the non-search part of the searching process (i.e. identification, recognition, decision, response execution, etc.).

Keywords: aviation security; X-ray baggage screening; search time; non-search time; training effect; visual search

Introduction

Threat detection using X-ray images in airport security screening is a process that only recently has become a major interest in research concerning object recognition and inspection. The task of airport security screeners is to recognise threat objects of various categories (guns, knives, improvised explosive devices (IEDs), etc.) in passenger baggage. By applying findings about object recognition, this extremely important part of common airport security concepts can be improved. The knowledge about how objects are perceived has allowed the creation of a computer-based training system X-Ray Tutor (XRT) (Schwaninger 2005). This training system considers the factors influencing the recognition of objects, which are the viewpoint an object is depicted in, the superposition by other objects in the bag and the number and type of other objects in the bag (Wallis and Bülthoff 1999, Schwaninger 2003, Schwaninger *et al.* 2004). XRT is individually adaptive. It starts with threat items depicted in easy views and increases image difficulty for each individual trainee by showing threat items in more difficult views and in more complex bags and with increasing superposition by other objects. In order to prevent screeners from memorising images of bags, combinations of images of bags and threat objects are created at the point of use. This approach considers the individual training level and visual-cognitive abilities of each screener.

Security inspection is a form of visual inspection but there exist few studies quantifying human performance of security screening (see Gale *et al.* 2000, 2005, McCarley *et al.* 2004, Schwaninger *et al.* 2004, Liu *et al.* 2006). The last two have shown that training increases the threat detection performance of airport security screeners dramatically. A deeper comprehension of the effect of training could be gained if the specific task of security inspection could be compared to more general models of industrial and other inspection tasks. Recent findings confirmed the applicability of a two-component model of visual inspection (Drury 1975, Spitz and Drury 1978) to X-ray screening data (Ghylin *et al.* 2006). Spitz and Drury (1978) assumed the inspection task was composed of search and decision components. Each of these components occupies part of the time needed for completing the task. Using the equations formulated by Drury (1975) and Spitz and Drury (1978) the total

*Corresponding author. Email: a.schwaninger@psychologie.uzh.ch

ISSN 0014-0139 print/ISSN 1366-5847 online

DOI: 10.1080/00140130802526935
http://www.informaworld.com

inspection time can be divided into the functional components of search and decision time.

The general model created and tested by Drury (1975) and Spitz and Drury (1978) is:

$$\text{P(detect a true target at or before time t)} = \left[1 - \exp\left(-\frac{(t - NST_{hit})}{ST_{hit}}\right)\right] * Pd_{hit} \quad (1)$$

$$\text{P(make a false alarm at or before time t)} = \left[1 - \exp\left(-\frac{(t - NST_{FA})}{ST_{FA}}\right)\right] * Pd_{FA} \quad (2)$$

where ST_{hit} or ST_{FA} = mean search time for hits or false alarms, NST_{hit} or NST_{FA} = mean non-search time for hits or false alarms, Pd_{hit} or Pd_{FA} = the probability of detection found from the raw data for hits or false alarms and t = the various raw reaction times (RTs) obtained from the data.

The model assumes an approximately exponential relationship between the time needed for searching a target and the cumulative probability of detecting a target (Morawski *et al.* 1980). Search time includes the visual scanning of an area to be searched (i.e. eye movements) and is terminated by either directing the attention to a suspicious part of this area (i.e. potential threat object in this case) or by deciding to stop searching. Decision time is everything except search and is more correctly called non-search time. It includes, among other things, the fixation of the suspicious object, the matching of the visual stimulus with representations stored in the visual memory, the decision (i.e. actually is threat object or not) and the time to execute the response. This model has been applied successfully to a number of different screening datasets (e.g. Ghylin *et al.* 2006). Applying this two-component model of visual inspection helps identifying the sub-processes of the whole inspection task and therefore may give some evidence about how the two processes improve differentially due to training.

Feature Integration Theory (FIT) (Treisman and Gelade 1980) assumes that visual features of objects are represented in feature maps. Features are those stimulus attributes that are processed rapidly and in parallel across the field of view. As soon as a visual field of an observer contains more than one object the binding problem arises (Treisman 1998). Features of various objects have to be combined correctly and assigned to the right object in order to perceive it correctly. In the original feature integration model (Treisman and Gelade 1980), search for feature conjunctions is not allowed. Wolfe (1994) found that a combination of feature information permits the efficient, guided search for feature conjunctions and postulated this in the Guided search 2.0 model of visual search. When a threat object is, probably deliberately, stowed in a bag, it is typically not just a target among several distractors. Most likely its shape on the X-ray image is interrupted by other objects surrounding it and superimposed on it. This complicates the assignment of features to an object, particularly if this object is only poorly known. If airport security screening training for threat object detection has the effect of creating internal representations of trained objects and storing them and making them available in the visual memory, then detection should improve because features are known and recognised better. It is also to be expected that, with growing knowledge about the visual appearance of threat objects in X-ray images of passenger bags, the number of required features for the recognition of an object can be limited or once separately perceived features can be combined as belonging to one object and thus becoming one feature. This would require building new feature maps. Considering the assumption of FIT that visual search for a combination of features is serial and therefore more time consuming than the visual search for a unique feature, the assumption would be that detection time would decrease for threat objects that are detected better. In other words, with increasing detection performance due to training the detection time, more explicitly the search time, should decrease.

Ghylin *et al.* (2006) found an enhancement of both the search process and the non-search process of inspection in the search for IEDs. This study extends the analysis to three other threat categories (guns, knives and other threat objects), potentially validating the findings of Ghylin *et al.* (2006) for objects other than IEDs. Prior inspection research has found individual differences among inspectors to be large (Czaja and Drury 1981a, Dollinger and Hoyer 1996, Wang *et al.* 1997, McPhee *et al.* 2004, Schwaninger *et al.* 2004, Riegelnig and Schwaninger 2006) so analyses of detection performance and RTs in the present study are controlled for age, gender and on-the-job experience.

There has been little examination of screener demographics in relation to either overall performance parameters or search and non-search measures. In the broader inspection literature, age, gender and experience have received some attention. Older inspectors tend to work more slowly and, at times, have lower detection performance (Czaja and Drury 1981a, McPhee *et al.* 2004), although any deficits can be largely negated by age-specific training (Czaja and

Drury 1981b). The mechanisms for age-related deficits are quite well understood (e.g. Fozard 1990). Ageing decreases pupil diameter, spatial resolution, visual acuity (particularly dynamic), contrast sensitivity (Owsley *et al.* 1983), depth perception and visual search (e.g. Plude and Hoyer 1986) but not colour vision or temporal resolution. Gender has not been found to be related to inspection performance (e.g. Wang and Drury 1989). Experience can either refer to novice/expert differences or to the effective length of experience of those with expertise. For example, Dollinger and Hoyer (1996) found novice/expert differences while Leach and Morris (1998) found no effect of longer experience. These findings are typical of experience results. Novices differ considerably from experts, but length of time on the job beyond initial training may not show much effect.

Methods

Participants

A total of 193 airport security screeners of a European airport, all with on-the-job experience of airport X-ray screening, were used. Of these 193 screeners, 98 (44 females, mean age 36.3 years, mean time on job 3.0 years; 54 males, mean age 40.0 years, mean time on job 3.0 years) of them trained for 6 months with XRT while the other 95 (48 females, mean age 35.1 years, mean time on job 3.0 years; 47 males, mean age 36.9 years, mean time on job 3.3 years) received no training with XRT during this period. All then took the X-Ray Competency Assessment Test (X-Ray CAT).

Materials and procedure

The X-Ray CAT is composed of 128 X-ray images of bags. Each image can include one threat object out of four threat categories according to international threat image projection system specification (guns, IEDs, knives and other threat objects). Stimuli were created from Smiths-Heimann Hi-Scan 6040i colour X-ray images. Each bag was used once containing a threat object and once containing no threat object, giving a total of 256 images. In each threat category, 16 objects were depicted once in an easy viewpoint (frontal view) and once in a more difficult rotated viewpoint (85° vertical or horizontal rotation). All threat objects were combined with bag images so as to ensure that for each threat object the degree of superposition (i.e. how much the threat object is superimposed by other objects in the bag) was equal for both viewpoints. The difficulty of the bags was equated across all categories and viewpoints by matching prior data on false alarm (FA) rates for each bag image. For a more detailed description of X-Ray CAT, see Koller and Schwaninger (2006).

The X-Ray CAT is a component of the XRT training system and can be programmed to appear anytime when a screener is training. The only visible difference for the screener between test and training was that feedback did not appear during the CAT test. The appearance of the images was the same for both training and test and therefore no instruction was needed. The images of the test disappeared after 10 s. Screeners had to decide whether the X-ray images contained a threat object or not (not OK or OK response). Difficulty ratings had to be provided by changing the position of a slider on a 100-point scale. Response times for each image were measured.

Results

The XRT program measured the response as hit, miss, false alarm or correct rejection with the RT for each test image and each participant. Detection performance in terms of A′ (Pollack and Norman 1964) was calculated for each threat category separately (see Figure 1), using the following formula (Grier 1971):

$A' = 0.5 + [(H - F)(1 + H - F)]/[4H(1 - F)]$, where H is the hit rate and F the false alarm rate. If the false alarm rate is greater than the hit rate, the equation must be modified ((Aaronson and Watt 1987): $A' = 0.5 - [(F - H)(1 + F - H)]/[4F(1 - H)]$. A′ scores were subjected to a univariate analysis of covariance (ANCOVA) with age and years on job as covariates, threat type treated as within-participants factor (guns, IEDs, knives, other) and gender and training (trained vs. untrained group) as between-participants factors. The results are summarised in Table 1.

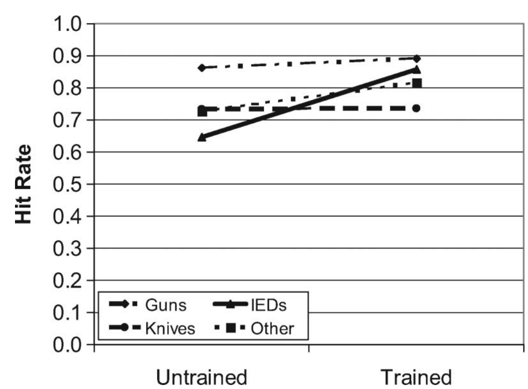

Figure 1. Detection performance A′ for trained and untrained participants for guns, improvised explosive devices (IEDs), knives and other threat objects. (Note: Performance values are multiplied by an arbitrary constant for security purposes).

Table 1. Results of univariate analyses of covariance on A′, hit rate (HR) and false alarm rate (FAR).

Factor	A′				HR				FAR			
	df	F	η^2	*p*	df	F	η^2	*p*	df	F	η^2	*p*
Threat Type (TT)	–	–	–	–	3, 561	7.12	0.04	< 0.001	NA	NA	NA	NA
Training (T)	1, 187	99.03	0.35	< 0.001	1, 187	31.51	0.14	< 0.001	1, 187	12.03	0.06	< 0.001
Gender (G)	1, 187	11.44	0.06	< 0.001	–	–	–	–	1, 187	6.25	0.03	< 0.05
T × TT	3, 561	52.55	0.22	< 0.001	3, 561	70.87	0.28	< 0.001	NA	NA	NA	NA
T × G	1, 187	3.99	0.02	< 0.05	–	–	–	–	–	–	–	–
TT × G	–	–	–	–	–	–	–	–	NA	NA	NA	NA
Age (A)	1, 187	40.25	0.18	< 0.001	1, 186	10.97	0.06	< 0.01	1, 187	4.97	0.03	< 0.05
Years on Job (Y)	1, 187	7.91	0.04	< 0.01	1, 187	7.31	0.04	< 0.01	1, 187	15.74	0.08	< 0.001
T × A	–	–	–	–	–	–	–	–	–	–	–	–
T × Y	–	–	–	–	–	–	–	–	–	–	–	–
TT × A	3, 561	7.21	0.04	< 0.001	3, 561	8.47	0.04	< 0.001	NA	NA	NA	NA
TT × Y	3, 561	5.16	0.03	< 0.01	3, 561	5.77	0.03	< 0.01	NA	NA	NA	NA

NA = not applicable.

A′ scores were also subjected to separate univariate ANCOVA for each threat (see Table 2).

Additionally, hit rates were subjected to a univariate ANCOVA with gender and training as between-participants factors, threat as within-participants factor and age and years on job as covariates (see Table 1 for details).

Separate ANCOVA on hit rates for each threat category were performed (see Table 2).

An ANCOVA on false alarm values with gender and training as between-participants factors and age and years on job as covariates shows a significant main effect of training (see Table 1 for details and values on covariate effects).

The effect of training in detection performance comprised an increase in the hit rate and a decrease in the false alarm rate. Figure 2 displays the hit rate for all threat categories for trained and untrained participants. The pattern is very similar to the one for the detection performance A′ (see Figure 1). This could be an indication that the difference of the detection performance between trained and untrained participants can mainly be attributed to a change in the hit rate. Nevertheless, the ANCOVA on false alarm values showed a significant effect of training, which means that also the false alarm rate is affected by training.

Average RT (seconds) was calculated for hits, false alarms, misses and correct rejections for trained and untrained screeners (see Figure 3). RTs were subjected to an ANCOVA with age and years on job as covariates, gender and training as between-participants factors and answer as a within-participants factor (hit, false alarm, miss, correct rejection). The results including the significance values can be seen in Table 3.

RTs were subjected to separate univariate ANCOVA for hits, false alarms, misses and correct rejections with gender and training as between-participants factors (detailed results are summarised in Table 4).

The intercorrelation matrix in Table 5 shows the correlations between the predictor variables (age, gender and years on job) and the performance variables (hit rate, false alarm rate, RTs of hits, false alarms, correct rejections and misses) as well as intercorrelations between the set of performance measures.

The mean search time and the mean non-search time were calculated for each screener individually and for hits and false alarms separately by applying the inspection model (Spitz and Drury 1978, see Introduction) to the RTs. If there were less than five responses (i.e. hit or false alarm) available for a person or a RT exceeded 14 s, these data were discarded from analysis. Data elimination was effected for 3.7% of the

Table 2. Results of univariate analyses of covariance on A′, hit rate and search and non-search time for the threat categories separately.

	Guns				Knives				IEDs				Other			
Factor	df	F	η^2	p	df	F	η^2	p	df	F	η^2	p	df	F	η^2	p
A′																
Training	1, 187	39.68	0.18	< 0.001	1, 187	24.42	0.12	< 0.001	1, 187	161.50	0.46	< 0.001	1, 187	51.38	0.22	< 0.001
Gender	1, 187	5.39	0.03	< 0.05	1, 187	5.74	0.03	< 0.05	1, 187	11.47	0.06	< 0.001	1, 187	7.37	0.04	< 0.01
Age	1, 187	20.01	0.10	< 0.001	1, 187	17.86	0.09	< 0.001	1, 187	37.05	0.17	< 0.001	1, 187	30.67	0.14	< 0.001
Years on Job	1, 187	13.01	0.07	< .001	1, 187	12.79	0.06	< 0.001	–	–	–	–	1, 187	7.93	0.04	< 0.01
Hit Rate																
Training	1, 187	6.37	0.03	< 0.05	–	–	–	–	1, 187	77.85	0.29	< 0.001	1, 187	29.45	0.14	< 0.001
Gender	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Age	1, 187	3.97	0.02	< 0.05	–	–	–	–	1, 187	14.35	0.07	< 0.001	1, 187	12.80	0.06	< 0.001
Years on Job	–	–	–	–	–	–	–	–	1, 187	10.37	0.05	< 0.01	1, 187	5.70	0.03	< 0.05
Search Time																
Training	–	–	–	–	–	–	–	–	1, 186	20.91	0.10	< 0.001	1, 187	4.33	0.02	< 0.05
Gender	1, 187	14.44	0.07	< 0.001	–	–	–	–	1, 186	5.43	0.03	< 0.05	1, 187	14.19	0.07	< 0.001
Age	1, 187	16.80	0.08	< 0.001	1, 187	12.55	0.06	< 0.001	1, 186	15.62	0.08	< 0.001	1, 187	19.41	0.09	< 0.001
Years on Job	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Non-Search Time																
Training	1, 187	48.60	0.21	< 0.001	1, 187	45.47	0.20	< 0.001	1, 186	66.75	0.26	< 0.001	1, 187	78.69	0.30	< 0.001
Gender	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Age	1, 187	9.04	0.05	< 0.01	1, 187	31.99	0.15	< 0.001	1, 186	32.15	0.15	< 0.001	1, 187	17.92	0.09	< 0.001
Years on Job	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

IED = improvised explosive device.

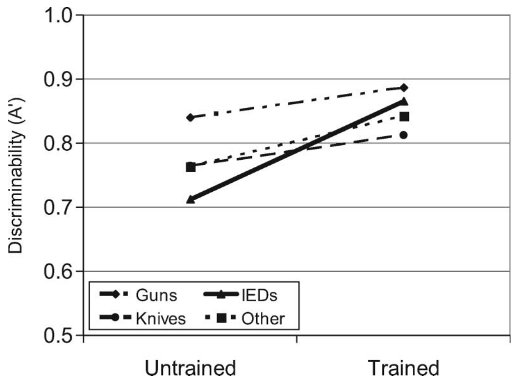

Figure 2. Hit rate for trained and untrained participants for guns, improvised explosive devices (IEDs), knives and other threat objects (Note: Performance values are multiplied by an arbitrary constant for security purposes).

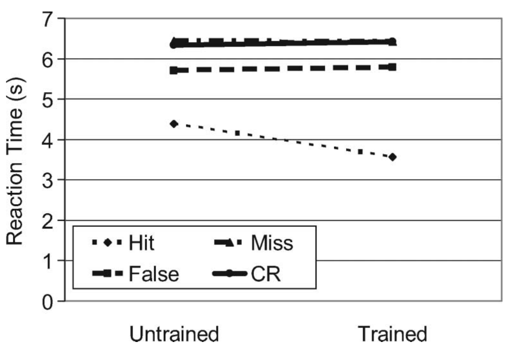

Figure 3. Reaction times (seconds) for trained and untrained participants for hits, misses, false alarms and correct rejections (CR).

trials (1839 of 49,407 cases with RT bigger than 14 s). Final sample sizes were n = 98 (hit RT) and 97 (false alarm RT) for trained participants and n = 95 (hit as well as false alarm RT) for untrained participants.

The hits were again separated by threat category (guns, IEDs, knives and other threat items), where for the IEDs one untrained participant achieved less than five hits within 14 s. Sample size therefore was 94 for IEDs. For false alarms, no threat categories exist.

The two-component model was fitted by first taking all of the response times in each threat category for each participant and ordering them from smallest to largest to form a cumulative distribution of response times. Because Pd_{hit} and Pd_{FA} were already known (they are the overall values of each probability) Equation (1) could be transformed to become:

$$\left[(1 - P(Detect))/Pd_{hit} \equiv \exp\left(-\frac{(t - NST_{hit})}{ST_{hit}}\right)\right]$$

Equation (2) is transformed in the same way. Taking natural logarithms of both sides gives a linear

Table 3. Results of univariate analyses of covariance on reaction time (RT).

Factor	RT df	F	η^2	p
Answer (S)	3, 561	15.51	0.08	< 0.001
Training (T)	–	–	–	–
T × S	3, 561	12.36	0.06	< 0.001
Age (A)	1, 187	18.87	0.09	< 0.001
Gender (G)	1, 187	11.90	0.06	< 0.001
Years on Job (Y)	–	–	–	–
T × A	–	–	–	–
T × G	–	–	–	–
T × Y	–	–	–	–
S × A	–	–	–	–
S × G	3, 561	3.72	0.02	< 0.05
S × Y	–	–	–	–

equation relating P(detect) to t, which can be fitted using linear regression (Drury 1975, Spitz and Drury 1978). This enables the parameters of search time (mean = ST_{hit}) and non-search time (mean = NST_{hit} to be estimated so that the separate functions can be differentiated mathematically. Again the calculations for false alarms are similar. Figures 4 and 5 show search and non-search times for false alarms and for hits per threat category, respectively.

The applicability of the two-component model (Spitz and Drury 1978) to security inspection task was tested by goodness of fit values (r^2) averaging above 0.9 both for hits (mean 0.961, SD 0.035) and false alarms (mean 0.905, SD 0.069).

The scores of search times and non-search times for hits were subjected to separate ANCOVA with age, gender and years on job as covariates, threat as a within-participants factor and training as a between-participants factor. Table 6 summarises the results.

Individual search and non-search parameters were subjected to separate univariate ANCOVA for each threat category, with training and gender as independent factors and age and years on job as covariates. Non-search time was affected by training in all threat categories (see Table 2 for details). The covariate effect of age was significant for all threat categories, whereas gender and years on job never had covariate effects (see Table 2). Table 2 also displays the effects on search time. Search time was only affected by gender for the categories of guns, IEDs and other threat objects, with small to medium effect sizes, but not for knives. Age had an effect on all threat categories and years on job on none. Training influenced only the search time for the categories of IEDs and other threat objects, but not for guns and knives.

Gender had significant effects on A′ (also on each threat category separately) and false alarm rate with

Table 4. Results of univariate analyses of covariance on reaction time for each answer category separately.

	Hits				Misses				False Alarms				Correct Rejections			
	df	F	η^2	*p*	df	F	η^2	*p*	df	F	η^2	*p*	df	F	η^2	*p*
Training	1, 187	38.05	0.17	<0.001	–	–	–	–	–	–	–	–	–	–	–	–
Age	1, 187	33.90	0.15	<0.001	1, 187	14.08	0.07	<0.001	1, 187	11.97	0.06	<0.001	1, 187	14.53	0.07	<0.001
Gender	1, 187	7.57	0.04	<0.01	1, 187	10.21	0.05	<0.01	1, 187	13.18	0.07	<0.001	1, 187	9.45	0.05	<0.01
Years on Job	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

small to medium effect sizes according to Cohen (1988; all analyses of effect sizes are interpreted according to Cohen 1988) but not on the hit rate. The effect of training is slightly influenced by gender, which means that there was a significant interaction between gender and training with a small effect size (see also Figure 6). There was a significant effect of gender on the RT with a medium effect size (see Tables 1–3).

Age had significant effects on A′ (also on each threat category separately), on the false alarm rate and on the hit rate (all threat categories except knives). The effect size was large for A′ and small to medium for hit and false alarm rates. Age also had a significant effect on the RT with a medium effect size (see Tables 1–3).

Job experience (years on job) had significant effects on A′ (all threat categories except IEDs) as well as on hit rate (only on IEDs and other threat objects) and false alarm rate. Effect sizes were small to medium. RT was not affected by working experience (see Tables 1–3).

Significant correlations were found between the detection performance A′ and age and between the detection performance A′ and years on job, separated for threat category (see Figures 7 and 8). However, if controlled for age, the correlation between detection performance A′ and job experience (years on job) is no longer significant ($r = 0.069$, $p = 0.34$).

Discussion

The data from this study show a substantial increase of threat detection performance in X-ray security screening due to training, regardless of age, gender or on-the-job experience, confirming prior findings of McCarley *et al.* (2004), Schwaninger and Hofer (2004) and also of Ghylin *et al.* (2006) in a more limited similar study of IEDs. Age and gender were taken into account regarding evidence that many aspects of cognition are impaired due to ageing (for a review see Craik and Salthouse 2000) and that gender influences cognitive tasks (see Halpern 1992 for a review). Although the present results point to such differences in detection performance with partly significant values, the effect sizes are small to medium according to the conventions of Cohen (1988). But, if so, detection performance decreases with increasing age and also with increasing job experience. However, since the correlation disappears with a partial correlation, controlling for age, this last interrelation presumably exists because age and job experience are confounded variables. As for gender, males perform overall slightly better than females. As mentioned before, detection performance increase due to training was not affected by these factors except for gender, where males tend to benefit a little more than females. Riegelnig and

Table 5. Intercorrelation matrix with predictor variables (gender, age, years on job (YOJ)) and performance variables.

	Correlations								
	pHit	Gender	pFA	RTHit	RTFA	RTCR	RTMiss	Age	YOJ
pHit									
Pearson Correlation	1	0.016	0.509**	−0.216**	0.028	0.282**	0.218**	−0.184*	0.168*
Sig. (2-tailed)		0.827	0.000	0.003	0.699	0.000	0.002	0.011	0.020
pFA									
Pearson Correlation	0.509**	0.153*	1	−0.007	−0.229**	0.130	0.084	0.086	0.266**
Sig. (2-tailed)	0.000	0.033		0.926	0.001	0.071	0.246	0.232	0.000
RTHit									
Pearson Correlation	−0.216**	−0.186**	−0.007	1	0.772**	0.695**	0.680**	0.323**	−0.004
Sig. (2-tailed)	0.003	0.009	0.926		0.000	0.000	0.000	0.000	0.953
RTFA									
Pearson Correlation	0.028	−0.281**	−0.229**	0.772**	1	0.828**	0.816**	0.263**	−0.068
Sig. (2-tailed)	0.699	0.000	0.001	0.000		0.000	0.000	0.000	0.347
RTCR									
Pearson Correlation	0.282**	−0.257**	0.130	0.695**	0.828**	1	0.936**	0.270**	0.041
Sig. (2-tailed)	0.000	0.000	0.071	0.000	0.000		0.000	0.000	0.572
RTMiss									
Pearson Correlation	0.218**	−0.259**	0.084	0.680**	0.816**	0.936**	1	0.252**	0.066
Sig. (2-tailed)	0.002	0.000	0.246	0.000	0.000	0.000		0.000	0.360
Age									
Pearson Correlation	−0.184*	−0.140	0.086	0.323**	0.263**	0.270**	0.252**	1	−0.045
Sig. (2-tailed)	0.011	0.052	0.232	0.000	0.000	0.000	0.000		0.531
YOJ									
Pearson Correlation	0.168*	−0.030	0.266**	−0.004	−0.068	0.041	0.066	−0.045	1
Sig. (2-tailed)	0.020	0.679	0.000	0.953	0.347	0.572	0.360	0.531	

*Correlation is significant at the 0.05 level (two-tailed).
**Correlation is significant at the 0.01 level (two-tailed).
FA = false alarm; RT = reaction time; CR = correct rejection.

iSchwaninger (2006) contributed a more detailed study on the influence of age and gender on detection performance in X-ray screening.

The aim of the present study was to analyse the effect of training on a more explicit level, i.e. which functions (search, decision) of the threat detection process change. High goodness of fit values ($r^2 > 0.90$) confirm the applicability of the two-component inspection model to X-ray security screening data, as was also found by Ghylin *et al.* (2006). The application of the two-component inspection model (Spitz and Drury 1978) to X-ray screening data allows a more detailed investigation of the effect of training on inspection performance of X-ray security screeners. Findings of industrial inspection studies on the cumulative distributions of RT for visual search showed that the inspection process can be divided conceptually and operationally into two sub-processes. The search process comprises the actual searching of an area (i.e. by a sequence of eye movements); the non-search process comprises all other components of the search (e.g. identification, recognition, decision, response execution, etc.). A similar model has been proposed by Gale *et al.* (2005) for use on inspection studies. This comprises an initial glance, then serial search followed by 'detection and interpretation'. Eye movements, rather than cumulative distribution fitting, were proposed to validate that model. Using the two-component inspection model, portions of the RT, i.e. the time needed for the whole search, can be assigned to one of the two sub-processes. Analysing search and non-search time can give a better understanding of which processes of search change due to training. These findings also quantify the effectiveness of the training system.

Comparing the RTs of trained and untrained screeners reveals a significant decrease for hits but not for false alarms, correct rejections or misses. This, and also the effect of training on hit rates, proves the effectiveness of the training system. Screeners have to learn to detect threat items and, therefore, hit rates increase and the time needed for detection decreases. Although RT for false alarms was not affected

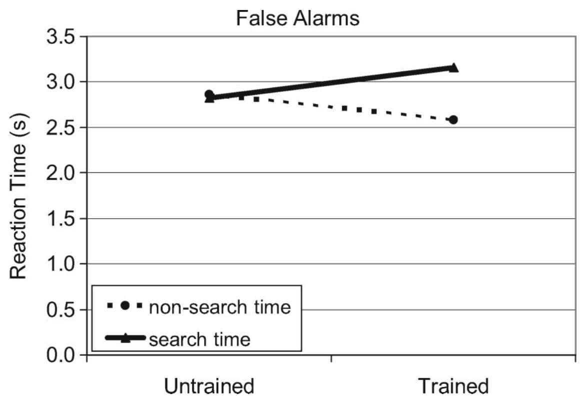

Figure 4. Search time and non-search time for false alarms by trained (n = 97) and untrained (n = 95) participants.

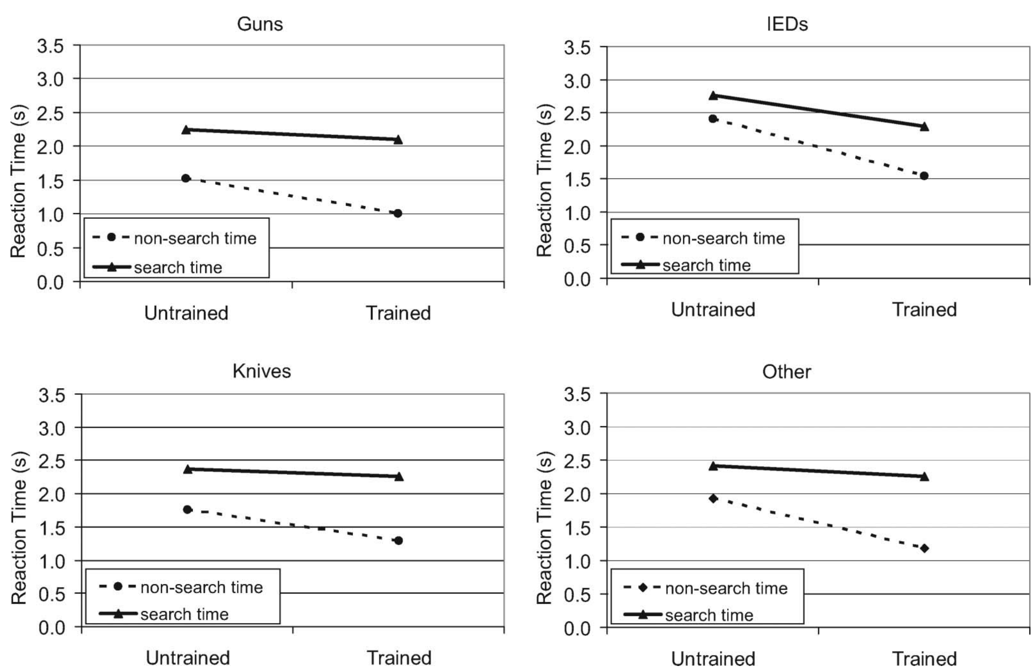

Figure 5. Search time and non-search time for hits, calculated separately for guns, improvised explosive devices (IEDs), knives and other threat items by trained (n = 98) and untrained (n = 95; IEDs n = 94) participants.

Table 6. Results of univariate analyses of covariance on search times and non-search times.

Factor	Search Times df	F	η^2	p	Non-search Times df	F	η^2	p
Threat Type (TT)	–	–	–	–	–	–	–	–
Training (T)	1, 186	8.59	0.04	<0.01	1, 186	83.38	0.31	<0.001
Gender (G)	1, 186	11.10	0.06	<0.01	–	–	–	–
T × TT	3, 558	5.93	0.03	<0.001	3, 558	11.95	0.06	<0.001
T × G	–	–	–	–	–	–	–	–
TT × G	3, 558	3.07	0.02	<0.05	–	–	–	–
Age (A)	1, 186	21.81	0.11	<0.001	1, 186	31.14	0.14	<0.001
Years on Job (Y)	–	–	–	–	–	–	–	–
T × A	–	–	–	–	–	–	–	–
T × Y	–	–	–	–	–	–	–	–
TT × A	–	–	–	–	3, 558	9.03	0.05	<0.001
TT × Y	–	–	–	–	–	–	–	–

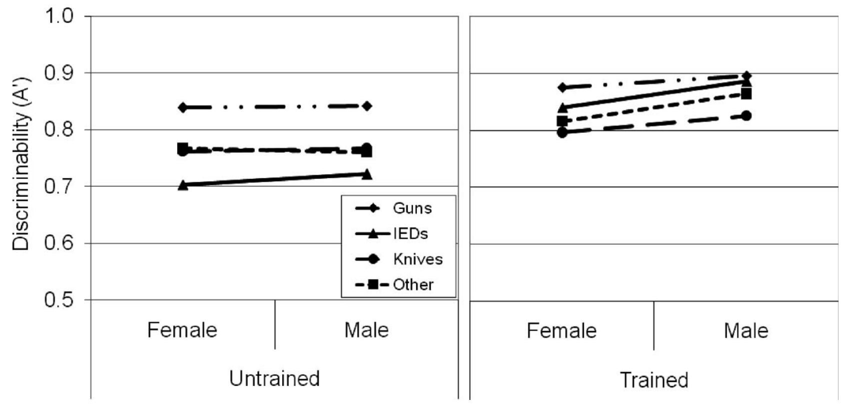

Figure 6. Detection performance A′ for trained and untrained participants for guns, improvised explosive devices (IEDs), knives and other threat objects for males and females separately. (Note: Performance values are multiplied by an arbitrary constant for security purposes).

significantly by training, false alarm rate actually was. However, as discussed in the next paragraph, RT of search and non-search components of false alarms both showed an effect of training.

The comparison of the search and the non-search time of trained and untrained screeners revealed a significant decrease of the non-search time for hits due to training, specifically to each threat category. Search time of hits was also affected by training. Examining the separate threat categories, this effect was only significant for IEDs and other threat objects, but not for guns and knives. Presumably guns and knives are well-known objects for experienced screeners so that training does not affect their detection substantially. For false alarms, the tendency for non-search time goes in the same direction but for search time the untrained screeners performed faster. These contrary effects lead to the fact that, overall, RT for false alarms shows no difference between trained and untrained screeners, as mentioned earlier. The substantial iincrease of detection performance A′ for trained screeners indicates the more effective search behaviour they achieve due to training. As trained screeners also showed significantly lower non-search times than untrained screeners, the training effect seems to come from faster detection, matching of ominous objects seen in the bag with memory representations of prohibited threat objects and faster recognition and decision about whether the object actually is a threat. Note also that the faster search times of trained screeners for hits were not accompanied by reduced times for correct rejection, i.e. the times when the screener failed to find a threat and moved to the next bag image. Thus, search was, in fact, more thorough after training, in that stopping time was a greater multiple of mean search time for hits. An improvement in speed of search is possible if threat objects are learned and therefore stored in the visual memory, which indicates that the training system effectively provides more exemplars of threat images. It

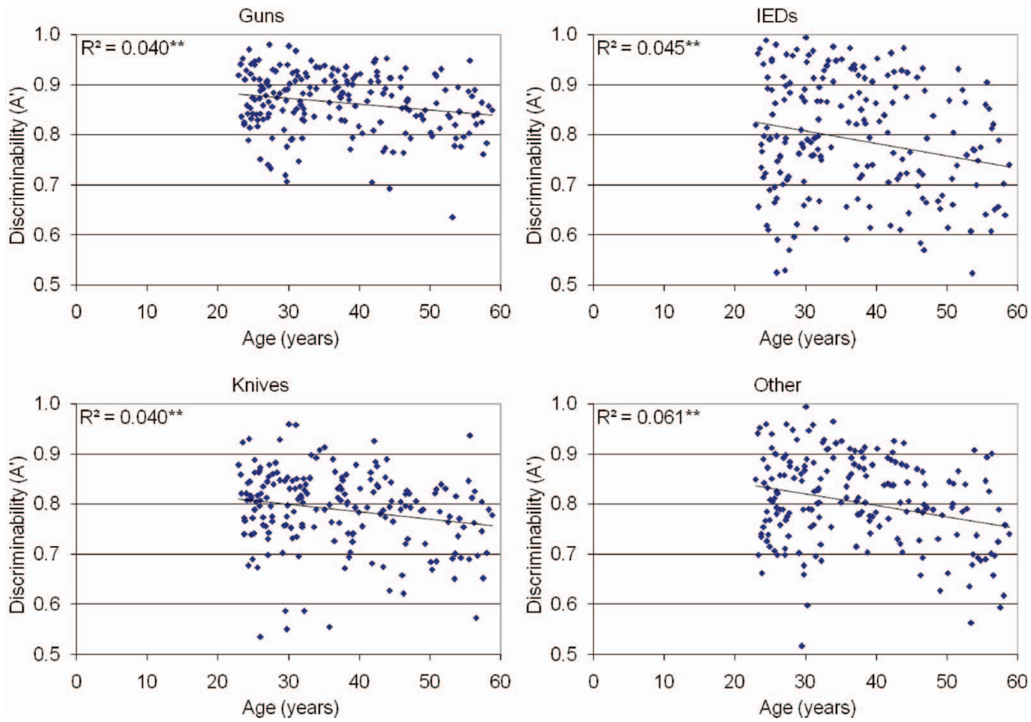

Figure 7. Correlation between age and A′ of each threat category. **Correlation is significant at the 0.01 level. IEDs = improvised explosive devices.

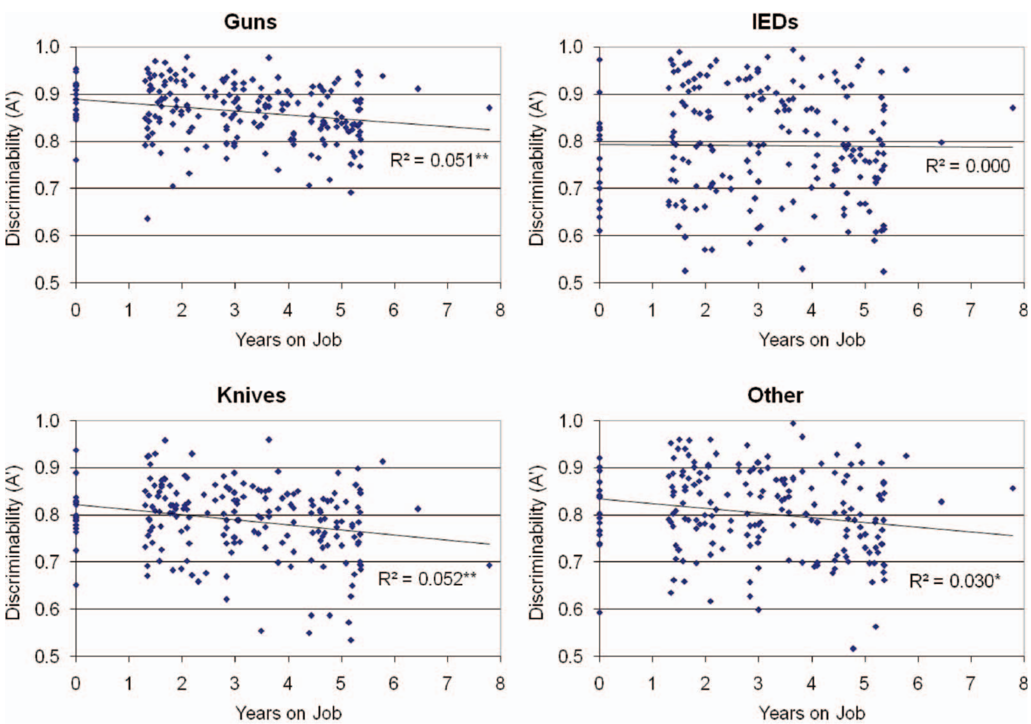

Figure 8. Correlation between years on job and A′ of each threat category. *Correlation is significant at the 0.05 level; **correlation is significant at the 0.01 level. IEDs = improvised explosive devices.

becomes easier to recognise common attributes of threat objects because they are represented in the visual memory.

A decrease of search time implies faster search, but only for IEDs, confirming the findings of Ghylin *et al.* (2006), and for other threat items. It remains to be investigated why search only gets faster for IEDs and slightly faster for other threat objects, but not for guns and knives. The former category has wide visual variability as IEDs can take many forms and their individual components can be scattered throughout the bag. In all studies of security X-ray inspection, they are harder to detect if untrained (Drury *et al.* 2006). In contrast the 'other' threat items list changes frequently so that it can never achieve the same level of familiarity as guns or knives, or even IEDs. Thus, IEDs have more 'head room' for improved detection and other threat items are always in a learning cycle. It has also to be observed if this speed-up continues with training or if it is a side effect of the dramatic improvement of detection performance of IEDs, which is far greater than for the other categories. This would imply that search stops improving as soon as detection performance of IEDs is at a higher level for all screeners. A possible reason for the decrease of search time for IEDs and other threat objects could be the building of new feature maps (Treisman and Gelade 1980) for IEDs and some threat objects belonging to the category of other (e.g. gas spray, taser, etc.). The assumption is that prior to training they exist with smaller probability than feature maps for guns or knives, because IEDs and other threat objects are rarely to never seen in everyday life, unlike guns or knives, and therefore are mostly unknown to untrained screeners.

McCarley *et al.* (2004) analysed eye movement data for the examination of visual performance in an X-ray image interpretation task. Simulating the baggage screening task of aviation security screening officers, participants had to detect a gun or a knife in an X-ray image of a bag. As mentioned earlier, they found improvements of sensitivity and response times due to training. Eye movement analysis revealed that the sensitivity increase resulted from changes in observers' ability to recognise target objects and not because the effectiveness of visual scanning changes.

The applicability of the two-component inspection model (Drury 1975, Spitz and Drury 1978) to X-ray inspection provides the possibility to investigate more closely the inspection process and its change due to training and to gain more knowledge about the individual components within inspection. This helps in improving the technologies, procedures and methods currently in use for X-ray security screening and therefore optimises the human–system interface.

Acknowledgements

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

References

Aaronson, D. and Watt, B., 1987. Extensions of Grier's computational formulas for A′ and B″ to below-chance performance. *Psychological Bulletin*, 102, 439–442.

Cohen, J., 1988. *Statistical power analysis for the behavioral sciences*. Hillsdale, NY: Erlbaum.

Craik, F.I.M. and Salthouse, T.A., 2000. *The handbook of aging and cognition*. Mahwah NJ: Erlbaum.

Czaja, S.J. and Drury, C.G., 1981a. Aging and pretraining in industrial inspection. *Human Factors*, 23 (4), 485–494.

Czaja, S.J. and Drury, C.G., 1981b. Training programs for inspection. *Human Factors I*, 23 (4), 473–484.

Dollinger, S.M. and Hoyer, W.J., 1996. Age and skill differences in the processing demands of visual inspection. *Applied Cognitive Psychology*, 10, 225–239.

Drury, C., 1975. Inspection of sheet metal materials: model and data. *Human Factors*, 17, 257–265.

Drury, C.G., Ghylin, K.M., and Holness, K., 2006. Error analysis and threat magnitude for carry-on bag inspection. *In*: *Proceedings of the 50th Annual Human Factors and Ergonomics Society meeting*, San Francisco, CA. Santa Monica, CA: HFES, 1189–1193.

Drury, C.G., Spencer, F.W., and Schurman, D., 1997. Measuring human detection performance in aircraft visual inspection. *In*: *Proceedings of the 41st annual Human Factors and Ergonomics Society meeting*, Albuquerque, NM. Santa Monica, CA: HFES, 304–308.

Fozard, J.L., 1990. Vision and hearing in aging. *In*: J.E. Birren, R.B. Sloane, and G.D. Cohen, eds. *Handbook of mental health and aging*, 3rd ed. San Diego, CA: Academic Press, 150–170.

Gale, A.G., Purdy, K., and Wooding, D., 2005. Designing out terrorism: human factors issues in airport baggage inspection. *In*: D. de Waard, K.A. Brookhuis, R. van Egmond, and Th. Boersema, eds. *Human factors in design, safety, and management*. Maastricht: Shaker, 63.

Gale, A.G., *et al.*, 2000. Is airport baggage inspection just another medical image? *Medical Imaging: Image Perception and Performance. Progress in Biomedical Optics and Imaging*, 1 (26), 184–192.

Ghylin, K.M., Drury, C.G., and Schwaninger, A., 2006. Two-component model of security inspection: application and findings. *16th world congress of ergonomics, IEA 2006*, Maastricht, The Netherlands, 10–14 July 2006.

Grier, J.B., 1971. Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, 75, 424–429.

Halpern, D.F., 1992. *Sex differences in cognitive abilities*, 2nd ed. Hillsdale NJ: Lawrence Erlbaum Associates.

Koller, S. and Schwaninger, A., 2006. Assessing x-ray image interpretation competency of airport security screeners. *In*: *Proceedings of the 2nd international conference on research in air transportation, ICRAT 2006*, Belgrade, Serbia and Montenegro, 24–28 June 2006, 399–402.

Leach, J. and Morris, P.E., 1998. Cognitive factors in the close visual and magnetic particle inspection of welds underwater. *Human Factors*, 40 (2), 187–194.

Liu, X., *et al.*, 2006. Is that a gun? The influence and features of bags and threat items on detection performance. *In*: P.D. Bust, ed. *Contemporary Ergonomics*. London: Taylor and Francis, 17–22.

McCarley, J.S., *et al.*, 2004. Visual skills in airport security screening. *Psychological Science*, 15, 302–306.

McPhee, L.C., *et al.*, 2004. Age differences in visual search for traffic signs during a simulated conversation. *Human Factors*, 46, 674–685.

Morawski, T., Drury, C.G., and Karwan, M.H., 1980. Predicting search performance for multiple targets. *Human Factors*, 22, 707–718.

Owsley, C., Sekuler, R., and Siemsen, D., 1983. Contrast sensitivity throughout adulthood. *Vision Research*, 23 (7), 689–700.

Plude, D.J. and Hoyer, W.J., 1986. Age and the selectivity of visual information processing. *Psychology and Aging*, 1 (1), 4–10.

Pollack, I. and Norman, D.A., 1964. A non-parametric analysis of recognition experiments. *Psychonomic Science*, 1, 125–126.

Riegelnig, J. and Schwaninger, A., 2006. The influence of age and gender on detection performance and the criterion in x-ray screening. *In*: *Proceedings of the 2nd international conference on research in air transportation, ICRAT 2006*, Belgrade, Serbia and Montenegro, 24–28 June 2006.

Schwaninger, A., 2003. Evaluation and selection of airport security screeners. *AIRPORT*, 2, 14–15.

Schwaninger, A., 2005. Increasing efficiency in airport security screening. *WIT Transaction on the Built Environment*, 82, 405–416.

Schwaninger, A., Hardmeier, D., and Hofer, F., 2004. Measuring visual abilities and visual knowledge of aviation security screeners. *IEEE ICCST Proceedings*, 38, 258–264.

Schwaninger, A. and Hofer, F., 2004. Evaluation of CBT for increasing threat detection performance in x-ray screening. *In*: K. Morgan and M.J. Spector, eds. *The Internet society: Advances in learning, commerce and security*. Southampton, UK: WIT Press, 147–156.

Spitz, G. and Drury, C.G., 1978. Inspection of sheet materials – test of model predictions. *Human Factors*, 20, 521–528.

Treisman, A., 1998. Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 353, 1295–1306.

Treisman, A. and Gelade, G., 1980. A feature integration theory of attention. *Cognitive Psychology*, 12, 97–136.

Wallis, G.M. and Bülthoff, H., 1999. Learning to recognize objects. *Trends in Cognitive Sciences*, 3, 22–31.

Wang, M.J. and Drury, C.G., 1989. A method of evaluating inspector's performance differences and job requirements. *Applied Ergonomics*, 20 (3), 181–190.

Wang, M.J.J., Lin, S.C., and Drury, C.G., 1997. Training for the strategy in visual search. *International Journal of Industrial Ergonomics*, 20, 101–108.

Wolfe, J., 1994. Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1, 202–238.