May 2018

Behavior Detection Visual Search Task Analysis Project

Visual Search Battery Report

Prepared for

Science and Technology Directorate
U.S. Department of Homeland Security

Prepared by

RTI International
3040 E. Cornwallis Road
Research Triangle Park, NC 27709

RTI Project Number 0214428.006.000

RTI Project Number
0214428.006.000

Behavior Detection Visual Search Task Analysis Project

Visual Search Battery Report

May 2018

Prepared for

Science and Technology Directorate
U.S. Department of Homeland Security

Prepared by

RTI International
3040 E. Cornwallis Road
Research Triangle Park, NC 27709

This research has been funded by the Department of Homeland Security (DHS) Science and Technology Directorate.

CONTENTS

FIGURES

Number **Page**

TABLES

Number **Page**

1. EXECUTIVE SUMMARY

1.1 Background

As part of its ongoing research with the Department of Homeland Security (DHS) Science and Technology Directorate (S&T) and the Transportation Security Administration (TSA), RTI International has conducted several studies to examine factors associated with visual search performance of TSA personnel. This project is a continuation of that line of research and aims to improve TSA's understanding of the critical skills needed to conduct behavior detection (BD). In addition, by identifying on-the-job (OTJ) performance metrics and conducting a validation study, this effort will provide empirical data regarding the predictive validity of a newly tailored BD visual search test battery and information regarding the validity of different individual characteristics that may be important in visual search. Specifically, the goal was to inform selection of officers who will be certified in BD or X-ray by examining whether there is a difference between those who are successful at BD visual search and those successful at X-ray visual search and by determining what officer characteristics are associated with good X-ray or BD performance.

1.2 Methodology

We collected data from 90 Behavior Detection Officers (BDOs) and 122 Transportation Security Officers (TSOs) who volunteered to participate in the two-hour study. Officers completed four visual search and BD tasks: a video-based passenger observation task, a simulation-based BD task, a luggage image search task, and a vigilance task. The video-based passenger observation task and simulation-based BD task were designed to assess officers' multitasking abilities, attention to detail, and selective and sustained attention. The vigilance task was used to assess BDOs' attention to detail, selective attention, and sustained attention. Finally, participants completed a brief demographic survey and nine individual characteristic assessments. These assessments included working memory, conscientiousness, need for cognition, boredom proneness, job boredom, attentional self-regulation, occupational self-efficacy, stress state, and coping for task stressors.

1.3 Results and Discussion

Data were analyzed using Pearson correlations, one-way analysis of variance, and regression analysis to identify linkages between individual characteristics and visual search, examine the association between luggage screening skill and BD skill, and determine whether performance on the battery of tasks was predictive of OTJ BD performance.

Results showed that several individual characteristics were related to BDOs' and TSOs' performance on the BD battery. For TSOs, higher working memory capacity was associated with more accurate performance on the video-based passenger observation task and vigilance task. For BDOs, higher working memory capacity was linked to higher accuracy

scores on the luggage image visual search task and performance on the vigilance task. Interestingly, working memory capacity played a significant role in how well TSOs performed on the BD-oriented task and how well BDOs performed on the TSO-oriented task. For BDOs, having greater working memory skills may have allowed officers to quickly and accurately memorize the set of target objects required in the luggage screening task and thus perform well on the screening task. Similarly, for TSOs, having a higher working memory capacity may have allowed these officers to quickly learn and remember the set of target behaviors used in the simulation-based BD task, which may have helped them perform better compared to TSOs with lower working memory capacity.

Results also showed that for TSOs and BDOs performance on the simulation-based BD task and luggage visual search task were significantly, positively correlated with each other. Specifically, high performance on the luggage screening task was associated with more accurate performance on the simulation-based BD task, more accurate performance on the video-based passenger observation task, and more accurate performance on the vigilance task. These results suggest that for both sets of officers, being good at one task was associated with being good at the other tasks. In general, TSOs scored higher on the luggage screening visual search and video-based passenger observation task compared to BDOs. A possible explanation for this is that BDOs had to quickly "block out" the cues they had been certified to look for and then quickly learn a new set of cues to perform the task. Results of regression analyses examining which individual characteristics explained the most variance in performance on the BD battery visual search tasks showed that for TSOs, working memory capacity was the best indicator of performance on the luggage image search task, vigilance performance was the best indicator of TSO performance on the battery's BD tasks. Regression analyses for BDOs found that vigilance performance was the best predictor of performance on the luggage image search task and the simulation-based BD task. BDO need for cognition, occupational self-efficacy, and working memory all emerged as significant predictors of video-based passenger observation task. Given the importance of sustained and selective attention skills in performing BD tasks and similar watch-keeping tasks, the repeated emergence of vigilance as a predictor is not surprising. This suggests that vigilance performance is a strong predictor of performance on threat detection and BD tasks and may offer predictive value.

Finally, this study found that performance on the vigilance task was the best predictor of OTJ BD performance. Need for cognition, low boredom proneness, low post-task distress, and low avoidance stress-coping all showed a relationship with higher OTJ BD performance ratings. Taken together, these findings indicate that officers with a need for mentally challenging work who can remain engaged in that work and can effectively cope with the stress brought on by the work (i.e., are vigilant and do not become bored or over stressed) are more likely to be successful at BD.

Future research should continue to focus on developing valid and reliable skills-based assessments that can be used to measure visual search skills. In addition to exploring new testing methods, researchers should examine new ways to quantify and measure BD performance.

1.4 Major Takeaways

- Individual characteristics were related to BDOs' and TSOs' performance on the BD battery. Working memory capacity played a significant role in how well TSOs performed on the BD-oriented task and how well BDOs performed on the TSO-oriented task.
- High performance on the luggage screening task was associated with more accurate performance on the simulation-based BD task, more accurate performance on the video-based passenger observation task, and more accurate performance on the vigilance task. These results suggest that for both sets of officers, being good at one task was associated with being good at the other tasks.
- Performance on the vigilance task was the best predictor of OTJ BD performance. Need for cognition, low boredom proneness, low post-task distress, and low avoidance stress-coping all showed a relationship with higher OTJ BD performance.

2. INTRODUCTION

Approximately 1.2 million people fly within the United States every day. To keep these passengers safe, the Transportation Security Administration (TSA) employs a multilayer security system to ensure that the traveling public and the nation's transportation systems are protected. The Behavior Detection (BD) program serves an essential function in this multilayered security approach. Unique from other security capabilities within the TSA security system, the BD program, which was previously reserved for Behavior Detection Officers (BDOs) but is now comprised of Transportation Security Officers (TSOs) as well, trains officers to identify anomalous behaviors by observing passengers and comparing what they see to an established behavioral baseline. The goal of the program is to identify high-risk travelers and subject them to additional screening.

Effectively executing the duties and tasks associated with BD requires a unique set of knowledge, skills, and abilities (KSAs). Officers must be able to sustain attention for prolonged periods of time, selectively attend to passengers while filtering out irrelevant information, detect behavioral indicators when they occur, and maintain situational awareness of the operational environment. In addition to these visual search–oriented KSAs, officers must demonstrate proficiency in certain nonvisual search skills. Understanding the tasks and responsibilities required of the BD job function, the skills required to successfully perform these duties, and how both components contribute to the mission of deterring threats and safeguarding the nation's security system is vital for developing tools and procedures that can be used to inform the selection, and training of BD-certified TSOs.

Since 2014, RTI has been conducting evidenced-based research to better understand the requirements of the BD function (RTI International 2015a, b, 2016, 2017b). The goal of this line of research has been to examine the visual search and threat detection skills of BDOs and to develop tests and assessments that can reliably and validly assess these skills.

Initial research conducted by RTI through the BDO Basic Visual Search contract with the Department of Homeland Security Science and Technology Directorate (DHS S&T) examined the visual search processes used by BDOs to determine whether their visual search success correlated with the same processes used by TSOs and whether additional or different indicators were useful for explaining BDO visual search success (Spain, Hedge, & Ladd, 2016). Direct comparisons showed no significant differences in search accuracy between BDOs and TSOs on the visual search task but that TSOs assessed images faster and more consistently than BDOs. Further results showed that search speed and search consistency explained a significant amount of variance in search accuracy for both BDOs and TSOs. In addition, spatial ability explained a significant amount of variance in visual search

performance of TSOs, whereas video-game play frequency explained a significant amount of variance in search accuracy of BDOs (Spain, Ladd, & Hedge, 2016).

The second study conducted under the BDO Basic Visual Search contract aimed to identify and refine visual search tests tailored to the BDO position and collect data from BDOs using these tasks. The outcomes of this project included three reports for DHS S&T: one that summarized the state of the science in BDO-relevant visual search (Spain & Ladd, 2015); a second that identified and outlined the duties and tasks performed by BDOs, the KSAs required for those tasks, and the general competencies associated with successful BDO performance (RTI International, 2015a); and a third that described assessment tools and approaches that could be used to measure the unique visual search constructs associated with the BDO position (RTI International, 2015b).

Furthermore, RTI conducted interviews with several BD experts to identify core constructs and assessments used by other agencies that rely on BD professions as part of their security postures. Using this knowledge, RTI developed a BDO visual search battery that included three tasks designed to assess core constructs associated with BDO visual search, and we collected data from BDOs using these assessments at a laboratory at a Category I (medium security) and a Category X (high security) airport. Results from this study showed evidence of convergent validity (i.e., the degree to which variables that should be related, are indeed related) for performance on the BDO visual search tasks and certain demographic factors that were correlated with performance on the BDO task battery (Spain, Ladd, & Hedge, 2016).

In summary, RTI has been actively involved in evidence-centered research examining traits and characteristics associated with strong visual search performance of TSA security personnel. The goal of this project was to continue this line of research by further refining and developing assessment tools tailored to the visual search requirements of the BD job function. Specifically, the goals were the following:

1. Refine and develop a set of performance-based visual search assessments and a battery of cognitive and non-cognitive predictors that can be used to reliably and validly assess the BD skills of TSA screening personnel.
2. Identify on-the-job (OTJ) performance metrics of BD that can be used to validly and reliably measure BD performance and skills.
3. Conduct a validation study using a sample of BDOs and TSOs from selected airports to examine linkages between BD visual search, X-ray visual search, individual characteristics, and OTJ BD performance.

This report describes the results of an empirical study that addressed these research goals and furthered DHS S&T's understanding of the visual search process required for successful BD performance. The study was guided by the following questions:

- Are individual differences in cognitive and non-cognitive characteristics, as measured through individual characteristics surveys, linked to performance on the BD battery?
- Is there an association between performance on the luggage screening and BD visual search tasks? That is, does being good at one type of visual search increase the likelihood that you will be good at another? If not, what characteristics predict luggage screening performance? What characteristics predict BD visual search?
- Is performance on the BD Battery predictive of OTJ BD performance?

In the following section we describe the research preparation activities we conducted to identify the tests and assessments tailored to the BDO position. Following this discussion, we provide an overview of the research methodology and review the results of the study. The report concludes with a discussion of the results in relation to the goals of the research project and directions for future research.

2.1 Research Preparation Activities Overview

2.1.1 Conducting a State-of-the-Science Literature Review on BDO Visual Search

To help identify the visual search requirements of BD, the first task we undertook was a systematic review of the literature in cognitive psychology as it relates to BD-relevant visual search. As previously noted, although TSOs rely on visual scanning of computer images, BD tasks require crowd-scanning behavior to identify appearance and behavior anomalies in the traveling public. Although both activities require officers to engage in visual search, the cues that are relevant to the search process and the skills needed to identify those cues may differ. The aim of the literature review was to determine the core visual search skills and abilities pertinent to BD tasks, to identify the behavioral cues that professionals search for, and to identify existing assessments that have been used to examine BD performance.

2.1.2 Reviewing BDO Job Requirements

In addition to reviewing the relevant scientific literature, it was important to understand the job characteristics and job requirements of the BDO position more clearly. As such, the second step we undertook was a thorough review of available position descriptions, job manuals, job analyses, and available training documents, materials, and programs for the BDO position. By reviewing these documents, we sought to better understand the processes involved in conducting BD and the behavioral and cognitive constructs required for successful performance.

2.1.3 Interviewing Subject Matter Experts

Following the review of the job-related information and pertinent scientific literature, we interviewed subject matter experts (SMEs) with a background in the type of visual search conducted by BDOs. The purpose of these interviews was to identify critical factors

potentially related to BDO performance and to identify assessments that may align with the KSAs pertinent to visual search requirements of BDOs.

2.1.4 Identifying Assessments Tailored BD Visual Search

Following the interviews with SMEs and using input from the state-of-the-science review and job requirements review, the final step in preparing for the study was to assemble a behavior detection battery comprised of assessments that could be tailored to examine BD visual search skills. We selected three tasks to measure BD visual threat detection performance. The first was a *vigilance task* that required participants to monitor a static display for a critical signal that occurred at infrequent intervals. This task was designed to assess several competencies BD, including attention to detail, selective attention, and sustained attention. The second task was a *passenger observation task*, which asked participants to watch a video clip of an airport security line and indicate the presence of several behaviors or appearance factors among passengers. This task was designed to measure several BD visual search constructs, including attention to detail, critical thinking, situation awareness, visual observation, and sustained attention. Finally, a *behavior detection task*, which also asked participants to indicate the presence of several behaviors among passengers but in a computer simulated environment, was chosen. This task was designed to assess an individual's attention to detail, situation awareness, visual observation, and memory, all of which are important aspects for BDOs in detecting changes to environmental baselines and remembering behaviors or appearance factors of suspicious passengers.

2.1.5 Pilot Study

Once the behavior detection battery was developed, RTI conducted a pilot test with 28 BDOs at a Category I airport and 14 TSA Headquarters staff to collect feedback on visual search task content, clarity of task instructions, and ease of interacting or responding to assessments. Specifically, we asked pilot participants to complete the visual search tasks and rate how representative the tasks were to actual BD scenarios, the degree to which the scenarios capture the types of visual search activities performed during BD, the fidelity of the response options, and the ease of making detection decisions.

Pilot participants were also given the opportunity to provide additional feedback to improve the assessment. Using this feedback, RTI refined the assessments to ensure the tasks adequately reflected the types of visual search and behavior detection activities conducted during BD and to ensure these skills were measured accurately.

3. RESEARCH METHODOLOGY

3.1 Participants

We collected data from 90 BDOs/BD-certified TSOs (hereafter referred to as BDOs) and 122 non-BD certified TSOs (hereafter referred to as TSOs) who volunteered to participate in the 2-hour study. Sixty-one participants were from a Category X (high security risk) airport in the South region of the United States, 33 were from a Category X airport in the Midwest, 55 were from a Category X airport in the Northeast, and 64 were from a Category X airport in the West. Participants were recruited by TSA staff at each airport. Because participants completed the study during their regularly scheduled shift they did not receive additional compensation. TSOs across the airports had an average tenure with TSA ofsevenyears, and BDOs had an average tenure with TSA of 10 years. This difference is expected given that BDOs start in the TSO position before applying for and being selected to become a BDO. **Table 1** contains an overview of participant counts and tenure by airport.

Table 1. Participant Counts and Tenure by Airport

			Current Position		
			non-BD certified TSO	BDO/BD certified TSO	Total
Airport	1	Count	31	30	61
		Average Tenure with TSA	7 years	10 years	9 years
	2	Count	15	18	33
		Average Tenure with TSA	10 years	11 years	11 years
	3	Count	32	23	55
		Average Tenure with TSA	7 years	10 years	8 years
	4	Count	44	19	63
		Average Tenure with TSA	5 years	9 years	6 years
Total		Count	122	90	212
		Average Tenure with TSA	7 years	10 years	8 years

3.2 Materials

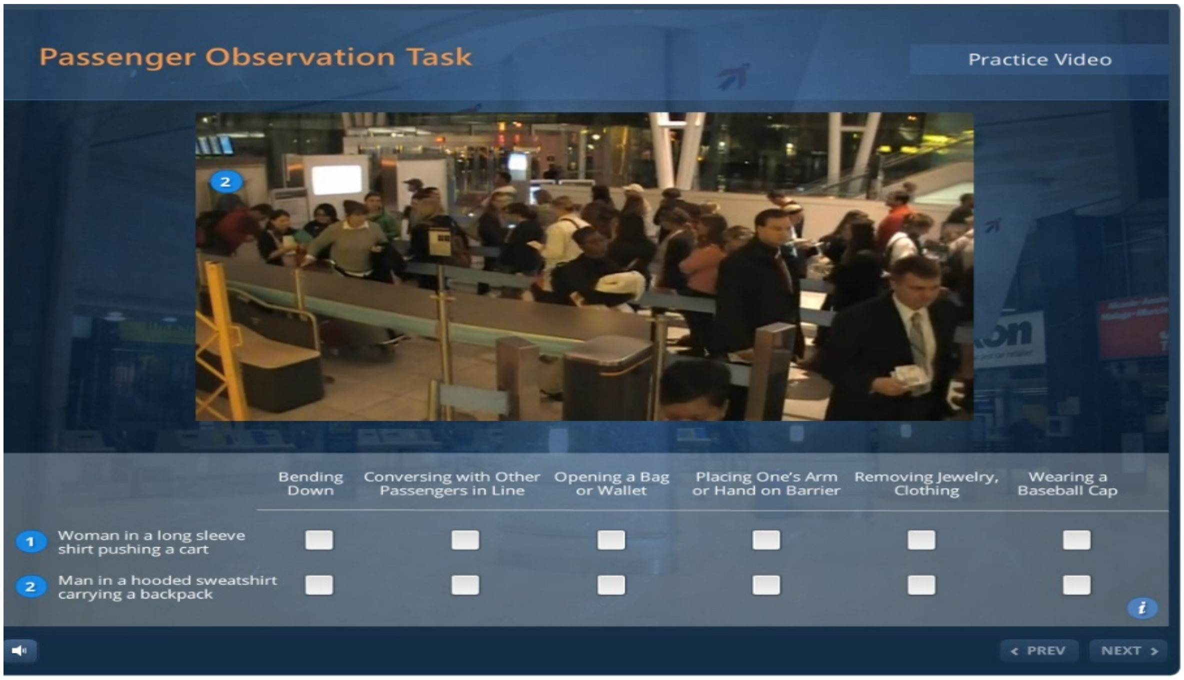
Participants completed a battery of BD tasks and a battery of surveys on a computer. The BD tasks were designed to assess the core visual search and visual observation skills

required of BD. The individual characteristics surveys contained items that assessed cognitive and non-cognitive traits, abilities, and preference hypothesized to be correlated with visual search performance.

3.2.1 Video-Based Passenger Observation Task

The video-based passenger observation task consisted of a five-minute video clip of airport passengers waiting in line at a security checkpoint and a list of associated behaviors and appearance factors. Test takers were asked to view the video clips and indicate, using the provided checklist, when a passenger displayed any of the selected indicators (see **Figure 1**). Previous versions of this task have demonstrated high levels of content validity (i.e., the content of the videos was representative of the visual search tasks performed by BDOs) and moderate levels of criterion-related validity (i.e., scores were correlated with an outcome measure), with empirical evidence showing positive correlations between test scores and BDO job performance ratings (Hendrickson et al., 2012). The task is designed to assess visual observation, memory, multitasking, and attention to detail skills.

Figure 1. Image of Video-Based Passenger Observation Task



The task was divided into two sections: an introduction section and a testing section. During the introduction section, participants learned about the task and the indicators they were required to detect. Specifically, participants were informed that they had to observe the passengers waiting in line and watch for six indicators (bending down; conversing with other passengers; wearing a baseball cap; opening a bag or wallet; placing an arm or hand

on barrier; or removing jewelry, clothing, or headphones). These indicators were different from the actual screening factors used by BDOs and were chosen based on item-level analyses and feedback from a pilot study (Hendrickson et al., 2012). Immediately following the task instructions, participants completed a brief practice session wherein they watched a 1-minute video clip and monitored the behavior of two designated passengers. Both passengers were identified in the video by a blue numbered circle that appeared over their heads and a tone that called the participants' attention to the video when it appeared. Participants placed a checkmark in the box next to any specific behaviors or appearance factors they observed.

After completing the practice session, participants advanced to the testing section wherein they watched a 5-minute video of passengers waiting in line at the checkpoint. Their job was to monitor the passengers and place a checkmark when a passenger displayed any of the selected indicators. Again, target passengers were identified in the video with a blue circle that appeared above their heads. In some instances, participants were required to monitor up to three passengers concurrently, making the task particularly difficult to perform. In total, there were nine times or instances in which a passenger displayed a target behavior or appearance factor during the video. Each correct detection was worth two points, which led to a total possible score of 18. Performance was calculated by dividing each participant's total score by the total possible score plus the number of false alarms (i.e., marking that a target behavior was displayed when it was not) they committed (see **Figure 2**). This equation penalized participants for committing false alarms by lowering the value of correct detections (by increasing the dominator) when individuals committed more false alarms.

Figure 2. Video-Based Passenger Observation Total Score Formula

$$Final\ Score = \frac{Score}{Total\ Possible\ Score\ +\ Number\ of\ false\ alarms} * 100$$

3.2.2 Simulation-Based Behavior Detection Task

In addition to completing the video-based passenger observation task, participants completed a simulation-based BD task (see **Figure 3**). This 12-minute exercise required participants to monitor synthetic passengers waiting in a security checkpoint and select any passengers who displayed any of the four target behaviors they were told to search for: wringing hands, clenching fists, patting chest, and checking their surroundings. Intermixed with these target cues were several non-target behaviors such as idle talking, checking watch, crossing arms, rubbing neck, and checking over shoulder. The purpose of including these distractor behaviors was to increase the realism and difficulty of the scenario. Participants were informed that when they saw someone engage in one of the target

behaviors to select the individual with their mouse and then report which behavior they displayed using the provided menu.

Figure 3. Image of Simulation-Based Behavior Detection Task

Over the course of the 12-minute simulation, participants monitored the behavior of approximately 150 passengers. There were 24 total target behaviors, which occurred at a rate of 2 per minute. Intermixed with these signals were 72 distracting behaviors. Participants received a final score at the end of the simulation, which included the overall number of correct detections.

A major difference between the simulation-based behavior detection task and the video-based passenger observation task was that for this task participants were not told which passengers to watch. Instead, participants had to actively scan all the passengers waiting in line and detect the presence of a target behavior. Another difference between the two tasks was that rather than relying on a checklist that listed the target behaviors, participants had to memorize the four target behaviors. Thus, the simulation-based BD task imposed more demands on working memory than the video-based task.

Assessing performance on the simulation-based behavior detection task. Prior to conducting analyses for the Behavior Detection Simulation, RTI considered several scoring approaches. The first awarded participants one point each time they correctly identified a target behavior

and then divided this sum by the total number of target behaviors (n = 24). Thus, the final score was operationally defined as the proportion of correct responses. One limitation with this measure is that it did not adequately consider other errors that participants could make, such as false alarms. The second scoring approach we considered awarded participants two points for a correct detection and divided this total by the sum of the total number of target behaviors displayed multiplied by 2, plus the number of false alarms committed by the individual. This equation therefore penalized participants for committing false alarms while also expanding the range of observed scores. One limitation with this scoring approach is that it did not account for instances in which participants mislabeled target behavior. Because the simulation required participants to select participants who displayed an indicator and then select the indicator the passenger exhibited from a list of options, participants could have correctly chosen a passenger who displayed one of the target behaviors but misidentified the behavior. The final scoring system that we considered accounted for these types of mislabeling errors. Performance on the task was assessed using the formula listed below:

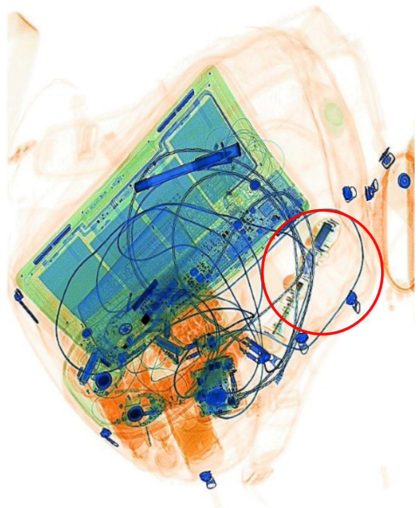
$$Score = \frac{\#\ of\ Correctly\ Identified\ Behaviors * 2}{(\#\ \text{of Hits} * 2) + (\#\ \text{of False Alarms}) + (\#\ \text{of mislabeled behaviors} * .5)}$$

Specifically, we calculated performance by dividing the number of behaviors correctly identified by sum of the total behaviors correctly identified (hits) plus the number of incorrect detections (false alarms) plus half of the total number of mislabeled behaviors. This scoring system penalizes participants for committing false alarms and for misclassifying behavioral cues.

3.2.3 Luggage Image Search Task

The third task included in the BD battery was a luggage screening visual search task. This task required participants to search for battery-operated improvised explosive devices (IEDs) in X-ray images of passenger luggage (see **Figure 4**). At the beginning of the task, passengers were shown images of five IEDs they had to detect in the trials that followed. Each IED was identified by the type of power source it used: either a 9-volt battery or two AA batteries; each power source had a distinct profile and shape. Then, participants completed 16 practice trials that preceded the experimental trials. During the practice trials, participants viewed luggage images and indicated if an IED was present or absent. Participants acknowledged the presence of an IED by selecting the "Stop Bag—IED Present" button on the display. Alternatively, participants acknowledged the absence of an IED by selecting the "Clear Bag" button on the display. After making this decision, participants used a 4-point rating scale to indicate their decision confidence (1—not at all confident; 2—slightly confident; 3—moderately confident; 4—fully confident). After submitting this rating, the computer provided participants with feedback about their decision accuracy.

Figure 4. Screen Shot of Luggage Image Search Task with IED Implanted



Following the practice trials, participants completed the experimental portion of the task. Participants viewed 100 luggage images; each image constituted a single trial (thus there were 100 trials). For each luggage image, participants determined if a target was present or absent by selecting the "Stop Bag—IED present" or the "Clear Bag—No IED present" buttons on the display. Then, participants rated their decision confidence. The purpose of including the decision rating scale is to differentiate between highly confident baggage stops and near misses. If no decision was made within the 20-second timeframe the image disappeared, and the next image appeared. The base rate of targets was set to 25 percent to simulate a lower prevalence rate compared to previous studies examining BDO and TSO visual search (Biggs, Cain, Clark, Darling, & Mitroff, 2013; Spain, Ladd, & Hedge, 2016) to be more reflective of "real-world" luggage screening prevalence rates but still be at a high enough rate to be able to detect meaningful differences in performance. At the end of the session, participants received automated feedback about their performance in the form of the overall number of hits, false alarms, misses, and correct detections they committed.

Performance on the luggage screening visual search task was assessed by examining participants' search accuracy, detection sensitivity, and detection bias. *Search accuracy* was operationally defined as the number of trials in which participants correctly identified the presence (hit) or absence (correct rejection) of a target out of 100 trials. Accuracy was calculated by summing the total number of hit and correct rejection trials (a score of 100 equals a perfect score). *Detection sensitivity* was calculated using the signal detection theory index *d′* (i.e., d prime), which quantifies an individual's ability to detect a "signal

event" while controlling for an individual's bias to respond during a decision-making task (see **Appendix A** for a computational definition; Green & Swets, 1966; Stanislaw & Todorov, 1999). Theoretically, *d′* scores range from 0 to infinity where a value of 0 indicates an inability to distinguish targets from noise during a decision-making task and large values indicate a correspondingly greater ability to distinguish targets from noise. In most situations, *d′* varies between 0.5 and 2.5 (See et al., 1995; Wickens & Holland, 2000).

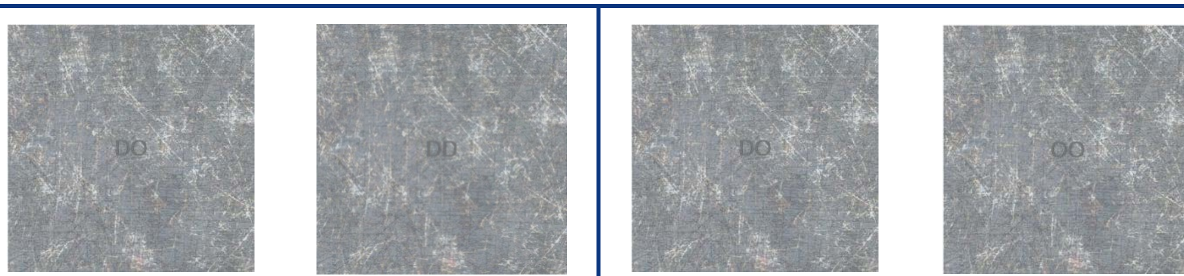
Response bias was calculated using the signal detection theory index *c*, which quantifies an individual's tendency to respond yes or no during a decision-making task. Scores were calculated by taking the average of the normalized z-scores that corresponded to the false alarm and hit rates and then multiplying the result by negative 1 (Stanislaw & Todorov, 1999). Negative values of *c* signify a bias toward responding "target present," whereas positive values signify a bias toward responding "target absent." If response bias is high the observer will be more likely to miss signals and will be less likely to make false alarms and will generally respond cautiously. If response bias is low, then observers will be more likely to respond, and errors will tend to take the form of false alarms rather than missed signals. A value of 0.0 is considered the neutral point (Stanislaw & Todorov, 1999). Detection sensitivity and detection bias were used in addition to search accuracy because they provided a means for identifying how well a TSO or BDO could find a luggage image with a target accounting for their tendency to respond.

3.2.4 Shortened Vigilance Task

Participants also completed a shortened vigilance task (SVT). This task was a modified version of the SVT developed by Temple et al. (2000) and required participants to monitor a computer display for a critical signal over a 10-minute watch period.

Critical signals were defined by the type of discrimination condition—simultaneous or successive. In the simultaneous condition, critical signals were identical pairs of letters (i.e., DD, OO, or two backward Ds). Mismatched pairs of letters (i.e., OD, DO) were considered neutral events and did not require a response. In the successive condition, the critical signal was OO only; all other letter combinations were neutral events. **Figure 5a** shows an illustration of a critical signal and neutral event presented against the mask during the simultaneous discrimination task. **Figure 5b** shows the critical signal in the successive condition. Condition assignment alternated across experiment sessions within airports (i.e., participants in session 1 received simultaneous condition, participants in session 2 received successive condition, participants in session 3 received simultaneous condition).

Figure 5. Illustration of Vigilance Stimuli



Neutral Event Critical Signal
(a) Simultaneous condition

Neutral Event Critical Signal
(b) Successive condition

The task was divided into five continuous 2-minute blocks. Within each block, the presentation order of the letter pairs was varied, with the restriction that the critical signal occurred with a probability of .20 (i.e., 24 times within each 2-minute block). Each letter pair was presented for 40ms against a cloudy gray visual mask. Letters were 8mm x 6mm in size, light gray in color, and constructed in 24-point Arial type font. Stimuli were presented at an event rate of 57.5 images per minute. This very high event rate was needed to simulate the psychophysical demands of longer vigilance tasks (Shaw et al., 2010).

Once a letter pair was presented, participants had approximately 1 second to respond by pressing the space bar. Correct responses made within this limit after the onset of a critical signal were recorded as hits. All other responses were categorized as misses, false alarms, or correct rejections. Participants were instructed to detect as many signals as possible but not to respond to non-signals so that the task is one requiring maximal performance.

Prior to starting the task, participants completed a 2-minute practice session during which a computerized female voice provided feedback on correct detections, misses, and false alarms; no feedback was provided during the main task. During the experimental session, participants complete the 10-minute task in either the simultaneous or successive condition (participants were randomly assigned to each condition). Participants did not receive feedback during the experimental portion of the task.

Performance was assessed by examining the proportion of correct detections, detection sensitivity (*d'*), and response bias (*c*) across trials. In addition to these three measures, we calculated vigilance decrement scores for each participant. Decrement scores were defined as the difference in response sensitivity from the first 2-minute block through the fifth 2-minute block. Decrements ranged from negative to positive values where negative values represented vigilance decrements and positive values represented increases in accuracy over the course of the task.

3.2.5 Summary of Constructs Assessed

As shown in **Table 2**, the assessments and testing tools included in the Behavior Detection Visual Search Battery assess multiple BDO visual search KSAs. The Video-Based Passenger Observation Task and Simulation-Based Behavior Detection Task closely resemble the BD tasks performed by BDOs on the job and measure attention to detail, sustained attention, selective attention, situation awareness, memory, and multitasking. The Vigilance Task and Luggage Image Search Task also assess competencies linked to the BDO position and should therefore provide meaningful information about BDO visual search capabilities.

Table 2. Competencies Assessed by Behavior Detection Visual Search Battery

Tasks	Attention to Detail	Situation Awareness	Sustained Attention	Selective Attention	Memory	Multitasking
Video-based Passenger Observation Task	X	X	X	X	X	X
Simulation-based Behavior Detection Task	X	X	X	X	X	X
Shortened Vigilance Task	X		X	X	X	
Luggage Image Search Task	X			X	X	

3.3 Individual Characteristics Measures

In addition to the BD battery, participants completed a battery of surveys designed to measure individual characteristics. The survey was delivered via the Web-based Evaluation Portal (WEP). Information collected from these surveys was used to examine correlates between individual characteristics and performance on the BD battery and performance on the job. A complete listing of the surveys is provided in **Table 3** (survey items are in **Appendix B**). A brief description of each measure follows. For a more detailed description of these measures see RTI International (2017a).

3.3.1 Demographic Questionnaire

The demographic questionnaire contains eighteen items and measures basic information from participants such as gender, age, race, ethnicity, tenure with TSA, tenure in current position, pay grade, level of education, and primary airport location. The demographic questionnaire was developed in house and was used to examine correlation between bio-data features and performance on the predictors and criteria.

Table 3. Surveys Included in Battery

Surveys	Survey Origin	Number of Items	Cronbach's Alpha*
Demographics	Created in-house	18	NA
OSPAN Working Memory Task	Unsworth et al. (2005)	75	NA
IPIP-NEO-Conscientiousness Scale	Goldberg et al. (2006)	20	0.93
Need for Cognition	Cacioppo & Petty (1982)	18	0.82
Boredom Proneness	Farmer & Sundberg (1986)	28	0.71
Job Boredom	Lee (1986)	17	0.94
Attentional Self-Regulation Scale	Luszczynska et al. (2004)	7	0.88
Occupational Self-Efficacy	Rigotti, Schyns, & Mohr (2008)	6	0.87
Short Stress State Questionnaire	Helton (2004)	24	
Engagement		8	0.82
Distress		8	0.85
Worry		8	0.79
Coping Inventory for Task Stressors	Matthews & Campbell (1998)	21	
Task Focused Coping		7	0.80
Emotion Focused Coping		7	0.84
Avoidance Focused Coping		7	0.64
Total		**234**	

* Cronbach's alpha is a score of internal consistency for assessing scale reliability. Scores range from 0 to 1 with higher scores reflecting more reliable measures. Cronbach's Alpha reported were calculated in this study.

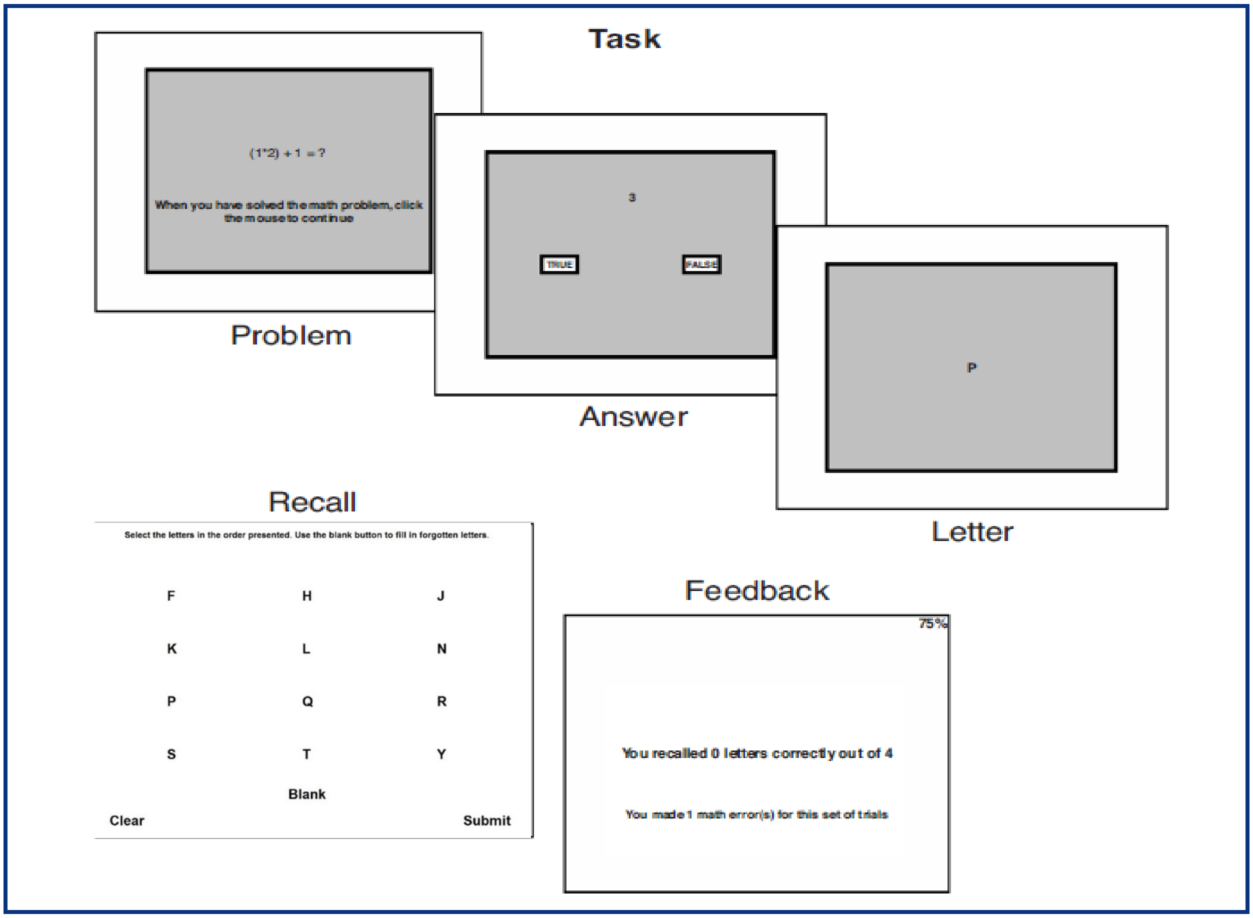
3.3.2 Working Memory Task

To test the memory capacity of officers, we used the Automated Operational Span Task (OSPAN), which is a computer-based measure of verbal working memory capacity (Unsworth, Heitz, Schrock, & Engle, 2005). The task requires participants to solve a series of math operations and memorize a set of unrelated letters before being asked to recall the letters in the correct order. The task consists of a practice section wherein participants become familiar with the different components of the task and an experimental section in which participants perform the actual task and data count toward their final scores.

The practice section contains three tasks: (1) a letter span task, (2) a math solutions task, and (3) a combined task. It is designed to provide targeted practice completing each task separately before the participant performs them together. During the letter span portion, participants practiced memorizing and recalling letter sets as they appeared on the computer screen. Letter sets contain three- to seven-letter strings, and each letter is presented for 800ms in the center of the screen. At the end of the letter set, participants

were presented with a 3 x 4 grid of letters and asked to recall the letters in the order in which they were presented. Participants made their responses by clicking on each letter (see illustration in **Figure 6**). After recall, the computer provided feedback about the number of letters correctly recalled in the current set.

Figure 6. Illustration of the O-SPAN Task



Next, participants practiced the math portion of the task, which served to familiarize them with the math operations they would need to solve. Participants first saw a math operation, which was a simple math equation (e.g., [3 − 1] + 2). Participants were instructed to solve the operation mentally as quickly as possible and then click the mouse to advance to the next screen. On the next screen, an answer to the math problem was presented and participants were required to click either a "true" or "false" box, depending on whether the answer was correct. Participants completed 15 practice math problems. Feedback was given on the screen after each math operation. The program captured the average time it took each participant to complete the math problems and used this time plus 2 standard deviations to establish a threshold for the testing portion of the task. If a participant

exceeded his or her average answering time plus 2 standard deviations during the testing phase of the task, the program automatically moved on to the next problem and counted that trial as an error.

In the final practice task, participants performed the letter recall and math portions together. Participants were first presented with the math operation. Once they solved the problem, and immediately after they indicated whether the math solution answer was true or false, participants were provided with a letter to be recalled. This sequence of math operator-answer-letter was repeated several times, based on the size of the letter sets. At the end of the set, participants were presented with a 3 x 4 grid of letters and asked to recall the letters in the order they were presented. Participants completed three rounds of the final practice session before progressing to the real trials.

The experimental trials were an extended version of the final practice session. For each trial, participants were presented with the math operation and were asked to solve it mentally as quickly as possible and to click their computer mouse to advance to the next screen. On the next screen, participants were required to decide whether the provided math answer was correct. Immediately following their answer, a letter was presented in the middle of the computer screen for 800ms. Following each complete set, participants were presented with a 3 x 4 grid of letters and asked to recall the letters in the order they were presented. Participants made their responses by clicking on the letters in the grid.

The number of letters and math problems presented to participants was based on the letter set size, which varies from three to seven. In a set size of three, participants viewed three math-answer-letter strings and were then presented with a 3 x 4 grid of letters and asked to recall the three letters in the order they were presented. In a set size of seven, respondents were presented with seven math-answer-letter strings and asked to recall the order of the seven letters presented. Participants complete each set size (i.e., 3, 4, 5, 6, and 7) three times for a total of 75 math operations and 75 letters. The order in which set sizes were presented was randomized. **Figure 6** provides an overview of the task sequencing. To ensure that participants did not trade off between solving the math operators and remembering the letters, we inspected the data for extreme scores on the recall portion and extreme misses on the math portion. No data were removed based on this screening procedure.

At the conclusion of the task, the program reported five scores: an overall span score, total number letters correct, math errors, speed errors, and accuracy. The overall span score, which is a measure of working memory capacity, was the sum of all perfectly recalled sets recalled by the participant. So, for example, if a participant correctly recalled three letters in a set size of three, four letters in a set size of four, and three letters in a set size of five, his or her OSPAN score would be (3+4+0) = 7. The total number correct was the total number of letters recalled in the correct position over the course of the experiment. The highest

score is 75. Speed errors reflect the number of times participants ran out of time in attempting to solve a math problem. Accuracy errors represent how many math problems were solved incorrectly. Finally, math errors represent the total number of speed errors and accuracy errors from the math task. The total number of letters correct was used for analytic purposes in this study.

3.3.3 IPIP-NEO-Conscientiousness Scale

Conscientiousness, or being purposeful, determined, and careful, was measured using 20 self-report items from the International Personality Item Pool (IPIP) NEO—Conscientiousness (NEO-C) scale (International Personality Item Pool, n.d.). Items are designed to measure several facets of conscientiousness including orderliness, dutifulness, and cautiousness (Goldberg et al., 2006). Participants reviewed each item and rated their level of agreement on a scale from 1 (very inaccurate) to 5 (very accurate) using a computer. Previous research shows that the scale correlates highly with the conscientiousness domain of the NEO (r = .88) and has high internal consistency (Cronbach's alpha =. 80). Rose et al. (2002) found a positive association between conscientiousness and vigilance performance but pointed toward the need to further investigate the role of this construct as a potential predictor of vigilance.

3.3.4 Need for Cognition

Need for cognition, or the extent to which a person looks for and enjoys engaging in mentally stimulating activities, was measured using an 18-item scale developed by Cacioppo and Petty (1992). Participants reviewed each statement and indicated how representative it was of themselves using a Likert scale of 1 (extremely uncharacteristic of me) to 5 (extremely characteristic of me). Because BD requires prolonged periods of observation and requires BDOs to engage in cognitively demanding activities while observing passengers, there could be a link between need for cognition and BD performance, such that those with a high need for cognition are better performers.

3.3.5 Boredom Proneness Scale

The Boredom Proneness Scale (BPS) is a 28-item instrument that measures trait-level sources of boredom. The scale was originally developed by Farmer and Sundberg (1986). Participants rated each item as it relates to themselves using a true or false scale. Following the theory that boredom leads to complacency and task disengagement, previous research (Sawin & Scerbo, 1995) has shown a negative association between trait boredom as measured by the BPS and performance in sustained attention tasks such as those performed by BDOs. Similarly, Kass, Vodanovich, and Callender (2001) found that individuals high in trait boredom are more prone to experience state boredom and perform more poorly on tasks requiring vigilance. Because the BD tasks associated with BDO

position require sustain attention, this scale was used to determine whether boredom proneness was negatively related to performance on the threat detection battery.

3.3.6 Job Boredom Scale

Created by Lee (1986), the Job Boredom Scale is a 17-item instrument arranged on a 5-point Likert Scale that measures state-level sources of boredom. Example items include, "Do you often get bored with your work?" and "Do you find the job dull?" Previous research has shown that employees with higher levels of tenure report higher overall boredom levels (Drory, 1982). Because officers engage in repetitive job tasks and because repeated exposure to the same stimuli (i.e., job tasks) leads to lower levels of arousal, job boredom could emerge as a significant predictor of performance on the BD battery.

3.3.7 Attentional Self-Regulation Scale

Attentional regulation was measured using seven items from the Self-Regulation Scale (Luszczynska, Diehl, Gutierrez-Dona, Kuusinen & Schwarzer, 2004). These items assess an individual's desire to accomplish his or her goals when presented with barriers and setbacks. Examples include "I can concentrate on one activity for a long time if necessary" and "If I am distracted from an activity, I don't have any problems coming back to the topic quickly." Participants rated each item using a scale of 1 (not at all true) to 4 (completely true). Luszczynska et al. (2004) found the scale had strong internal consistency (Cronbach's alpha coefficient of .77). Because BDOs are required to regulate their attention while performing BD duties, attentional regulation may emerge as a predictor of performance on the BD battery and as a construct that is related to job performance.

3.3.8 Occupational Self-Efficacy

Occupational self-efficacy was measured using the 6-item scale developed by Rigotti et al. (2008). Example items include "I can remain calm when facing difficulties in my job because I can rely on my abilities" and "When confronted with a problem in my job, I can usually find several solutions." Items were rated on a six-level response scale ranging from 1 (not at all true) to 6 (completely true). High values reflect high occupational self-efficacy and indicate the person feels they have the needed ability to successfully do his or her job (Rigotti et al., 2008). Research has shown it to be a mediator of stress (Grau, Salanove, & Peiró, 2001) and a predictor of job performance (Stajkovic & Luthans, 1998).

In addition, we added an item to measure mission commitment, which asks participants how committed they are to TSA's mission. The purpose of adding this item is to determine whether mission commitment correlates with performance on any of the tasks.

3.3.9 Short Stress State Questionnaire

The Short Stress State Questionnaire (SSSQ) is a 24-item measure designed to assess subjective stress states associated with task performance. It includes items related to

worry, engagement, and distress (Helton, 2004). Immediately following SVT task performance, participants completed each item using a self-reported scale from 0 (not at all) to 4 (extremely). Research by Matthews, Warm, Shaw, and Finomore (2014) has shown that stress states, as measured by the SSSQ, are strong predictors of vigilance.

3.3.10 Coping Inventory for Task Stressors

Finally, the Coping Inventory for Task Stressors (CITS) is a 21-item state-based measure of coping to stressful tasks (Matthews & Campbell, 1998). The inventory includes three subscales that measure task-, emotion-, and avoidance-focused coping. Items on each subscale are rated using a Likert scale from 0 (not at all) to 4 (very much). Recent research has shown that coping, as measured by the CITS, can predict vigilance performance better than personality measures (Matthews et al., 2014).

3.4 Criterion Measures of Interest

Another important aspect of this study was collecting measures of job performance to serve as criteria. These measures allowed the research team to examine associations between job performance and performance on the BD battery. This section contains information on how RTI collected criterion measures (i.e., job performance data) for the validation study. Criteria consisted of scores from the Performance Evaluation Checklist (PEC), which is part of BDOs' Annual Proficiency Review (APR) and ratings of BDO BD performance using a scale developed by RTI.

3.4.1 PEC Scores

RTI worked with TSA's Office of Security Operations (OSO) after data collection to receive PEC scores for each BDO scheduled to participate in on-site data collection. A more detailed review of the PEC scores is provided in the BDO Job Performance Measures Report (see RTI International, 2017b). Although not all scheduled BDOs completed the test battery, we obtained scores for all scheduled BDOs so as not to potentially reveal the identity of officers who either did not attend their scheduled session or declined to participate. Overall PEC ratings were obtained from the two most recent PEC evaluations, when available.

3.4.2 Job Performance Rating Scale

We also used a newly developed rating scale to evaluate visual search skills and BD performance. This scale, called the BDO Job Performance Rating Scale (JPRS), was developed by RTI and was completed by BDO supervisors. The following section describes the BDO JPRS and procedures used to collect ratings from supervisors.

Scale Overview

The JPRS was designed to measure technical aspects of visual attention and BD (see **Appendix C**). It contains 32 items clustered around five dimensions: Visual Observation

Skills (domain specific), Visual Observation Skills (domain general), Behavior Detection and Analysis, General Skills and Abilities, and Technical Performance. Item statements were developed by reviewing literature examining predictors of visual search performance in applied domains, including IED detection, and by reviewing BDO job task analyses and previously established behavioral anchors used in the PEC rating guide. Each item describes behaviors associated with exceptional visual search performance during BD. After developing a draft set of items, RTI gathered feedback from subject matter experts at TSA's OSO and from BDO experts. These individuals reviewed the items and their associated performance dimensions, edited wording and language, and provided suggestions on which items should be added or removed. Based on this feedback, RTI revised several of the items and finalized the scale.

The current version of the JPRS requires the supervisor to read each statement and then rate using a 4-point scale ranging from 1 (not well at all) to 4 (extremely well) how well the statement describes the team member's ability. Example items include "Maintains focus and attention over the course of BDA operations," "Focuses on relevant passengers' behaviors while ignoring other distracting information," "Concentrates his/her attention on a task for long periods of time," "Pays attention to detail," and "Accurately adjusts the environmental baseline to accommodate all changes in typical behaviors and appearances."

In addition to providing ratings for these BD-oriented skills and abilities, supervisors indicated how often in the past month they had an opportunity to observe the listed BDO they rated using a scale from 1 (almost never had the opportunity) to 5 (very frequently had the opportunity).

The JPRS should offer advantages over the current annual review ratings used by TSA (i.e., TOPS) because the items are designed to be more relevant to measuring visual observation skills and performance. The internal consistency values for each scale dimension and the overall JPRS scale are reported in **Table 4.**

Table 4. Internal Consistency of JPRS Scale

Dimensions	Number of Items	Cronbach's Alpha
Domain-Specific Visual Search Skills	6	0.94
Domain-General Visual Search Skills	8	0.95
Behavior Detection and Analysis	8	0.94
General Skills and Abilities	6	0.86
Technical Performance	2	0.91
Observation Frequency	2	0.97
JPRS Whole	32	0.97

4. PROCEDURE

4.1 Behavior Detection Battery Sessions

Upon arrival at the testing location, participants received two identical copies of an informed consent sheet from the RTI data collection staff. The informed consent sheet contained a description of the study, participant requirements, and information about voluntary participation and withdrawal. Officers who elected to participate in the study signed and dated one copy of the informed consent and returned that signed copy to the RTI data collection staff. The other, non-signed copy was for the officers to keep for their records if they so desired. Participants who declined to participate in the study were requested to stay in the visual search laboratory and given access to computer-lab computers to complete training or other job-related materials to review during their scheduled testing session. This was done to comply with Institutional Review Board regulations and ensure voluntary participation.

After completing the informed consent requirements, RTI staff logged each participant into the WEP on an RTI-owned laptop computer using a unique username and password (usernames and password were created for each participant prior to data collection using the information provided to RTI by the TSA airport point of contact). RTI staff briefly reviewed the experimental tasks with each participant and showed them how to use the WEP to complete the assessments. The WEP was used to sequence the experimental tasks such that participants would complete the tasks in the same order. **Table 5** outlines the order in which the assessments were completed, and the amount of time generally required for each. Participants were given the opportunity to ask questions about the study prior to completing the first task and at any point throughout the session. After completing all tasks participants were debriefed and released from the study.

4.2 JPRS Ratings

Prior to the data collection period at each airport, TSA staff provided the name of the supervisor of record for BDOs who were scheduled to participate in the study. RTI used this information to create customized, supervisor-specific JPRS documents that listed only the BDOs each supervisor needed to rate. These documents included JPRS forms with participant IDs and an employee-participant ID key needed to decipher which participant ID belonged to which BDO. We did not list BDO names directly on the JPRS form to avoid having BDO names and performance ratings on the same document.

Table 5. Study Procedure

Assessment Activity	General Time Estimate
1. Introduction and Demographic Survey	5 minutes
2. Working Memory Task (O-SPAN)	25 minutes
3. Short break	5 minutes
4. Simulation-Based Behavior Detection Task	15 minutes
5. Video-Based Passenger Observation Task	10 minutes
6. Luggage Image Search Task	15 minutes
7. Short break	5 minutes
8. Individual Characteristic Surveys	20 minutes
9. SVT (simultaneous or successive)	15 minutes
10. SSSQ and CITS	5 minutes
Total Time	**2 hours 0 minutes**

During the data collection period at each airport supervisors reported to the testing room at a time that was convenient for them. The supervisor and an onsite RTI staff member would then relocate to a nearby, private location out hearing and visual range of the testing location so the RTI staff member could provide an overview of the study and conduct a rater training with the supervisor. The overview included a discussion of the purpose of the research, the role and importance of the supervisor's ratings, the process for developing the scale, and an in-depth review of how to complete the rating form. The overview also informed raters that their ratings would be completely confidential and would be used for research purposes only, not for annual performance review cycles.

After the overview discussion, supervisors were given a rater error training that focused on halo error, allowing non-performance related factors influence ratings (e.g., family background, education), and same-level-of-effectiveness error. Halo error occurs when raters allow a general good or bad impression of the rate influence their job performance ratings. Same-level-of-effectiveness error occurs when raters tend to give the exact same rating to all rates. See **Appendix D** for the complete rater training.

Following the rater training, supervisors were asked to complete the ratings on their own at that time and to return the ratings to RTI staff upon completion. Some supervisors, however, were not able to complete the ratings immediately following the training due to workload demands and requested to take the forms with them to complete on their own time. These supervisors were requested to and agreed to complete the ratings in a confidential manner. After completing their ratings, supervisors who took the rating forms with them returned the forms to RTI staff before the end of the data collection period.

4.3 Planned Analyses

The primary objectives of this study were to (1) investigate whether individual differences in cognitive and non-cognitive characteristics were linked to performance on the behavior detection battery; (2) determine whether there was an association between performance on the luggage screening task and BD tasks and examine differences in performance on the BD task between BDOs and TSOs; and (3) identify whether performance on the BD battery was predictive of OTJ performance. To explore these questions, data were analyzed using a series of one-way analysis of variance (ANOVA) models using position type (BDO, TSO) as independent variables and performance metrics from the BD battery as dependent variables. All ANOVAs were tested with the p value of 0.05 unless otherwise noted. Correlations were also computed, and data were subjected to regression analyses to identify a set of factors related to BD performance for BDOs and TSOs. Data were analyzed using SPSS version 25.

Table 6 contains the analyses that were used to answer each research question.

Table 6. General Analysis Plan

General Research Questions	Analyses Conducted
R1. Are individual differences in cognitive and non-cognitive characteristics, as measured through individual characteristics surveys, linked to performance on the behavior detection battery?	R1. Computed correlations between individual characteristics and performance on the BD tasks for TSOs and BDOs.
R2. Is there an association between performance on the Luggage Image Visual Search and BD tasks? Does being good at one make you good at the other? If not, what characteristics predict Luggage Image Visual Search performance? What characteristics predict BD visual search?	R2. Computed correlations between performance on the Luggage Image Search Task, Simulation-Based Behavior Detection Task, and Video-Based Passenger Observation Task. R2. Regression examining predictors of performance on Luggage Image Search Task. R2. Regression examining predictors of performance on Simulation-Based Behavior Detection Task and Video-Based Passenger Observation Task.
R3. Is performance on the Behavior Detection Battery predictive of OTJ BD performance?	R3. Computed correlations between tasks in BD battery and JPRS ratings (for BDOs only).

5. RESULTS

5.1 Data Cleaning and Filtering

BD Battery Data

Prior to conducting any statistical analyses, data from the BD battery were inspected and filtered according to specific criteria. For survey measures that required a sum total score, cases were examined for item missingness and potential exclusion from the sum total score calculation if participants did not answer all of the items in the item set. No cases were excluded from sum total score calculations. Eight cases were identified as having individual characteristic scores exceeded a commonly used outlier threshold of three standard deviations above or below the group mean (Tabachnick & Fidell, 2001). These cases violated the threshold across multiple individual characteristic surveys indicating that they did not engage with the content of the measures and may have intentionally given the same answer for all items in an assessment (i.e., straight-lined). These eight cases were therefore excluded from the analytical sample. On the visual search tasks, five cases were identified for not participating in a given task. For example, they let the task play out without attempting to answer correctly. The final dataset included 204 participants. Descriptive statistics for each task and the demographic survey are provided in ***Appendix E***.

Although our overall sample size did not meet the threshold laid out in our pre-data collection power analysis, and therefore was not large enough to detect small differences in performance, it was still sufficiently large enough to detect significant differences in performance. Notably, however, because of the unknown a priori effects of clustering in our sample, wherein statistical power is reduced because officers employed at a particular airport are more similar than officers across airports, our resulting effective sample size may have been too small to detect a meaningful relationship between variables. See **Appendix F** for power analysis details.

OTJ BD Performance Data

Upon evaluation of the PEC scores received we determined that there was not enough variance in scores across officers for the data to be of use in validating the visual search battery. Of a maximum score of 39, officer PEC scores had a mean of 38.8 and standard deviation of 0.37. Scores for all officers ranged from 38 to 39. This lack of variance did not allow for detecting meaningful differences in officer ability as rated by the PEC. We therefore exclude an examination of PEC scores in the below results.

JPRS rating data demonstrated sufficient variance for analysis. We also examined JPRS rating data for outliers using the same three standard deviation threshold used for battery data. No cases were eliminated during this analysis. Lastly, we examined JPRS Observation Frequency data to determine if supervisors had sufficient opportunity to observe ratees and

would therefore be able to accurately rate officer performance. Our analysis showed that supervisors had sufficient opportunity to observe ratees ($M = 3.32$, $SD = 1.14$).

5.1.1 Research Question 1: Are individual differences in cognitive and non-cognitive characteristics, as measured through individual characteristics surveys, linked to performance on the behavior detection battery?

Our first research question asked whether individual-level differences in cognitive and non-cognitive characteristics are linked to performance on the BD battery. To answer this question, we computed and compared bivariate Pearson's correlation coefficients between the individual characteristics survey and BD visual search tasks. These comparisons were done separately for TSOs and BDOs. **Table 7** provides an abridged summary of the results; only visual search tasks with significant correlations to the listed individual characteristic are listed in the table. See **Appendix G** for the complete version of the table. Correlation values can range from -1.0 to +1.0, where 0 represents no relationship, -1 represents a perfect negative correlation, and +1 represents a perfect positive correlation. A significant positive correlation indicates that both variables change in the same direction, for example as one variable increases so too does the other variable. Conversely, a significant negative correlation indicates that the variables move in opposite directions, for example as one variable increases the other decreases. Note, a significant correlation does not mean that one variable causes the other to increase or decrease; it simply shows a relationship between two variables.

Summary of Correlations for TSOs

As shown in **Table 7**, there were both positive and negative statistically significant linkages between individual characteristics and performance on the BD battery for TSOs. This suggests that there were several individual characteristic factors related to TSO visual search performance. Working memory as measured by OSPAN showed the strongest positive, significant correlations, specifically with accuracy, $r(36) = .61$, $p < .001$, and detection sensitivity, $r(37) = .63$, $p < .001$, on the SVT—Successive task. This would suggest, as working memory scores increased, so too did signal detection accuracy and sensitivity (i.e., an individual's ability to detect a "signal" within a distraction-filled environment). Results showed that working memory also had a moderate linkage with scores on the video-based passenger observation task, $r(78) = .34$, $p = .002$.

CITS—Task-Focused coping scores had moderate, significant linkages to luggage visual search performance accuracy, $r(77) = .31$, $p = .007$, and luggage visual search detection sensitivity, $r(77) = .35$, $p = .002$. It had a slightly weaker significant link with SVT—Simultaneous detection sensitivity, $r(41) = .36$, $p = .022$. This indicates that officers who are able to maintain a task focus when in a stressful situation, rather than having an

emotionally negative reaction or trying to avoid the stressful situation, tend to be more accurate during luggage visual search activities.

Table 7. Correlations between Individual Characteristics and Behavior Detection Visual Search Performance for TSOs and BDOs (abridged)

Measures of Comparison	TSOs	BDOs
OSPAN—Working Memory compared to		
Video-Based Passenger Observation Task	r(76) = .34**	ns
Luggage Image Search Task *Accuracy*	ns	r(64) = .24*
Shortened Vigilance Task—Simultaneous Condition *Decrement*	ns	r(28) = .42*
Shortened Vigilance Task—Successive Condition *Accuracy*	r(34) = .61***	r(26) = .48*
Shortened Vigilance Task—Successive Condition *Detection Sensitivity*	r(35) = .63***	r(29) = .36*
Conscientiousness compared to		
Luggage Image Search Task *Detection Sensitivity*	r(78) = -.24*	ns
Need for Cognition compared to		
Simulation-Based Behavior Detection Task	ns	r(64) = .32**
Luggage Image Search Task *Accuracy*	ns	r(64) = .25*
Shortened Vigilance Task—Successive Condition *Accuracy*	ns	r(26) = .47*
Boredom Proneness compared to		
Simulation-Based Behavior Detection Task	ns	r(64) = -.37**
Job Boredom compared to		
Luggage Image Search Task *Accuracy*	r(77) = .22*	ns
Shortened Vigilance Task—Simultaneous Condition *Response Bias*	r(39) = .33*	ns
Attentional Self-Regulation compared to		
Shortened Vigilance Task—Successive Condition *Accuracy*	ns	r(26) = .42*
Commitment to TSA Mission compared to		
Shortened Vigilance Task—Successive Condition *Accuracy*	r(34) = .39*	ns
CITS—Task-Focused compared to		
Simulation-Based Behavior Detection Task	ns	r(62) = .33*
Luggage Image Search Task *Accuracy*	r(75) = .31**	ns
Luggage Image Search Task *Detection Sensitivity*	r(75) = .35**	ns
Shortened Vigilance Task—Simultaneous Condition *Detection Sensitivity*	r(39) = .36*	ns
Shortened Vigilance Task—Simultaneous Condition *Response Bias*	ns	r(29) = .44*
CITS—Emotion-Focused compared to		
Video-Based Passenger Observation Task	r(76) = -.30**	ns
Luggage Image Search Task *Response Bias*	r(75) = .24*	ns
Shortened Vigilance Task—Successive Condition *Accuracy*	r(33) = -.37*	ns
CITS—Avoidance-Focused compared to		
Shortened Vigilance Task—Simultaneous Condition *Accuracy*	r(39) = -.34*	ns

Measures of Comparison	TSOs	BDOs
Shortened Vigilance Task—Simultaneous Condition *Detection Sensitivity*	r(39) = -.32*	ns
Shortened Vigilance Task—Successive Condition *Decrement*	r(33) = .36*	ns

* p < .05; ** p < .01; *** p < .001; ns = not statistically significant

The results also showed that conscientiousness, job boredom, commitment to TSA's mission, CITS—Emotion Focused, and CITS—Avoidance Focused were related to BD battery performance. Specifically, results showed that as job boredom scores increased, so too did SVT—Simultaneous response bias (i.e., an individual's tendency to respond yes or no during a decision-making task where negative values indicate bias toward responding "target present" and positive values signify a bias toward responding "target absent"). This relationship suggests that as officer job boredom increases, that officer tends to report that no search target is present. For TSOs, this indicates that officers who find their job dull or boring may have a tendency to indicate no threat is present during luggage image search activities, likely due to inattention or lack of job task focus.

It is important to recognize that, even when correlations are statistically significant, the strength of the relationship between many of the variables is fairly low. The strength of the relationship (i.e., effect size) for a bivariate correlation is calculated by squaring the *r* value. For example, the strength of the relationship between job boredom and luggage image search task accuracy ($0.22^2 = 0.05$) shows that job boredom only explained 5 percent of the total variance in luggage image search task accuracy. To put these results in perspective, assume that all the factors imaginable (personality traits, demographic factors, differences in ability, etc.) can be used to predict 100 percent of the variability in a task. The current results show that job boredom explains only 5 percent of the variance in luggage image search task accuracy. The remaining 95 percent of variance in luggage image search task accuracy is explained by other factors.

Some demographic variables were also significantly correlated with BD battery performance (see **Appendix H**). Age and tenure with TSA were significant negative correlates with numerous battery tasks; meaning that as age and tenure increased, scores on the battery decreased. This could suggest an experiment mode effect where the mode of testing, namely a laptop computer, could have impacted scores. Another explanation is that older officers and officers with more tenure are in more supervisory, administrative roles and do not employ their visual search skills daily like younger officers. Relatedly, the final demographic item worth noting is that frequency of performing X-ray image search on the job was positively correlated with luggage image search task scores.

Summary of Correlations for BDOs

Results showed a similar pattern of linkages for BDOs as for TSOs (see **Table 8**). Working memory as measured by OSPAN showed the most linkages with BD battery performance. Specifically, it was positively correlated with SVT—Successive accuracy, $r(28) = .48$, $p = .011$, SVT—Successive detection sensitivity, $r(31) = .36$, $p = .046$, SVT—Simultaneous decrement, $r(30) = .42$, $p = .022$, and luggage image search accuracy, $r(66) = .24$, $p = .048$, although these correlations were not as strong as they were with TSOs. This suggests that officers who are better able to hold in working memory the objects they are searching for performed better on the vigilance and luggage search tasks. Interestingly, need for cognition showed several linkages for BDOs but not for TSOs. Need for cognition

Table 8. Correlations between Individual Characteristics and Behavior Detection Visual Search Performance for TSOs and BDOs Controlling for Age

Measures of Comparison	TSOs	BDOs
O-SPAN—Working Memory compared to		
Video-Based Passenger Observation Task	r(73) = .31**	ns
Shortened Vigilance Task—Simultaneous Condition *Decrement*	ns	r(25) = .48**
Shortened Vigilance Task—Successive Condition *Detection Sensitivity*	r(32) = .62***	ns
Need for Cognition compared to		
Simulation-Based Behavior Detection Task	ns	r(61) = .29*
Boredom Proneness compared to		
Simulation-Based Behavior Detection Task	ns	r(61) = -.38**
CITS—Task-Focused compared to		
Simulation-Based Behavior Detection Task	ns	r(59) = .31*
Luggage Image Search Task *Detection Sensitivity*	r(72) = .25*	ns
Shortened Vigilance Task—Simultaneous Condition *Detection Sensitivity*	ns	r(26) = .44*
CITS—Emotion-Focused compared to		
Video-Based Passenger Observation Task	r(73) = -.28*	ns
Luggage Image Search Task *Response Bias*	r(72) = -.28*	ns
CITS—Avoidance-Focused compared to		
Shortened Vigilance Task—Simultaneous Condition *Accuracy*	r(36) = -.33*	ns
Shortened Vigilance Task—Simultaneous Condition *Detection Sensitivity*	r(36) = -.34*	ns
Shortened Vigilance Task—Successive Condition *Decrement*	r(30) = .37*	ns

* $p < .05$; ** $p < .01$; *** $p < .001$; ns = not statistically significant

was positively correlated with simulation-based BD task, r(66) = .32, p = .010, luggage image search accuracy, r(66) = .25, p = .040, and SVT—Successive accuracy, r(28) = .47, p = .013, indicating that officers who look for and enjoy engaging in mentally stimulating activities performed better on these tasks.

The strongest linkage for BDOs was a negative link between boredom proneness and performance on the simulation-based BD task, r(66) = -.37, p = .003. This suggests that BDOs who are less prone to boredom were better able to recognize target behaviors over the duration of this 12-minute work sample task. As with TSOs, there were also significant correlations between BDO demographics and BD battery performance (see **Appendix F**).

Most notably, age and TSA tenure again had the most and strongest negative correlations with battery performance.

Exploratory Analysis of Linkages between Individual Characteristics and Behavior Detection Battery Performance Controlling for Age

Given the significant negative correlations between age and several of the visual search battery tasks, we repeated the above correlational analyses while controlling for the effects of age. Results are presented in **Table 8**. The analysis did not reveal any new significant linkages between individual characteristics and visual search tasks. It did, however, reduce the number of significant correlations between the two. When controlling for age, only working memory, CITS—Task-Focused, CITS—Emotion-Focused, and CITS—Avoidance-Focused remained as significantly correlated with visual search for TSOs. For BDOs, only working memory, need for cognition, boredom proneness, and CITS—Task-Focused remained. This suggests that the age of the officer completing the computer-based battery did indeed have an effect on his or her performance. What is not known, however, is whether factors not examined in this study are the actual cause of this difference. For example, as mentioned above, it could be that younger officers have more experience playing computer or video games and therefore have developed a level of skill at computer based, game-like tasks. On the other hand, it could be that as age and tenure on the job increase, these more experienced officers found the battery tasks boring and were not as engaged in the tasks as younger officers.

5.1.2 Research Question 2: Is there an association between performance on the luggage screening and behavior detection visual search tasks? Does being good at one make you good at the other?

Our second research question asked whether there was an association between performance on the luggage screening and BD visual tasks. To answer this question, we computed bivariate Pearson correlation coefficients between each of the tasks in the BD battery. A positive correlation between the tasks would suggest that the visual search skills required for successful luggage screening are linked to the watch-keeping skills required for BD analysis. **Tables 9** and **10** provide a summary of the results for BDOs and TSOs, respectively. Positive values in the table indicate that both variables increased linearly (both variables changed in the same direction), whereas negative values indicate that one variable increased as the second decreased (both variables move in opposite directions).

As shown in **Table 9**, performance on the luggage image task was positively correlated with performance on the behavior battery for TSOs. Specifically, the table shows that higher accuracy scores and higher detection sensitivity scores on the luggage detection task associated with higher scores on the simulation-based BD task, video-based passenger observation task, and the SVT. Positive correlations were also found between the

Table 9. Correlations on Behavior Detection Battery for TSOs

		1	2	3	4	5	6	7	8
1.	Luggage Image Search Task *Accuracy*	1							
2.	Luggage Image Search Task *Detection Sensitivity*	0.86**	1						
3.	Simulation-Based Behavior Detection Task	0.29**	0.38**	1					
4.	Video-Based Passenger Observation Task	0.25*	0.24*	0.37**	1				
5.	SVT—Simultaneous *Detection Sensitivity*	0.54**	0.58**	0.46**	0.37*	1			
6.	SVT—Simultaneous *Decrement*	0.20	0.08	0.03	-0.17	-0.06	1		
7.	SVT—Successive *Detection Sensitivity*	0.13	0.35*	0.40*	0.26	—	—	1	
8.	SVT—Successive *Decrement*	-0.23	-0.34	-0.05	-0.07	—	—	-0.06	1

* p < .05; ** p < .01

Table 10. Correlations on the Behavior Detection Battery for BDOs

		1	2	3	4	5	6	7	8
1.	Luggage Image Search Task *Accuracy*	1							
2.	Luggage Image Search Task *Detection Sensitivity*	0.88**	1						
3.	Simulation-Based Behavior Detection Task	0.47**	0.53**	1					
4.	Video-Based Passenger Observation Task	0.31*	0.34**	0.40**	1				
5.	SVT—Simultaneous *Detection Sensitivity*	0.22	0.41*	0.37*	0.55**	1			
6.	SVT—Simultaneous *Decrement*	-0.03	-0.08	-0.04	-0.12	-0.08	1		
7.	SVT—Successive *Detection Sensitivity*	0.48**	0.43*	0.42*	0.43*	—	—	1	
8.	SVT—Successive *Decrement*	-0.22	-0.39	-0.06	0.21	—	—	-0.20	1

* p < .05; ** p < .01

simulation-based BD, video-based passenger observation, and vigilance tasks, such that high performance on the simulation-based BD task was linked to high performance on the video-based passenger observation task and simultaneous version of the SVT. These results suggest that TSOs who were able to identify threats in luggage images were also able to identify behavior cues and appearance factors in the simulation-based BD-oriented tasks in the battery.

For BDOs, a similar pattern of results emerged (see **Table 10**). Specifically, positive correlations were observed between performance on the luggage image visual search, simulation-based BD, video-based passenger observation, and vigilance tasks. These findings suggest that BDOs who performed well on the BD-oriented tasks also performed well on the luggage image and vigilance tasks.

Overall these results suggest that the tasks within the battery measured similar constructs and provide empirical evidence that increased competence one task is linked to increased competence on the other BD and luggage image visual search tasks. Generally speaking, being good at one task was associated with being good at the others.

Next, we conducted a series of one-way ANOVAs to examine whether there were any performance differences on the BD battery between BDOs and TSOs.

Table 11 contains statistical comparisons between the two positions. Results showed that TSOs were significantly more accurate at identifying targets in the luggage image search task compared to BDOs and demonstrated higher levels of perceptual sensitivity. Overall, TSOs were 5 percent more accurate (77.20 vs. 73.25) and 20 percent more sensitive (1.79 vs. 1.49) than BDOs in identifying threats in the luggage image search task. These results are not surprising given that many of the BDOs who participated in the study had not been certified in luggage screening for several years. We also examined another metric that is commonly used in visual search research—response time, which measured how long it took participants to search an image and decide if a target was present or absent—but analyses failed to show a significant difference between BDOs and TSOs.

There was a non-significant difference between BDOs and TSOs on the Video-Based Passenger Observation Task; however, an inspection of the means shows that TSOs ($M = 60.81$) performed slightly better compared to BDOs ($M = 56.70$). Further analysis suggests that the effect size of the difference between the group means was small ($d = 0.25$) and that this comparison would have benefited from a larger sample size. Having a larger sample size may have allowed for finding a larger, significant difference between the two positions, or, conversely, would allow for confirmation of no meaningful difference between the two positions. It is also important to note that several BD-certified officers commented that it was hard to search for the set of proxy indicators, rather than the real indicators they have been certified to detect. A plausible explanation is that some officers experienced proactive interference, which is a cognitive phenomenon that occurs when old

information prevents the recall of new information (Kane & Engle, 2000). In this case, the trained set of indicators that BDOs are certified to find may have inhibited their ability to recall or detect the proxy set of behaviors required for the task. The performance differences observed between the two

Table 11. Means, Standard Deviations, and Results of the One-way ANOVAs Comparing Performance on Behavior Detection Battery

	TSO		BDO		ANOVA
Performance Variable	**Mean**	**SD**	**Mean**	**SD**	***F***
Luggage Image Search Task *Accuracy*	77.20	11.10	73.25	13.14	$F(1,146) = 3.95^*$
Luggage Image Search Task *Detection Sensitivity*	1.79	0.76	1.49	0.83	$F(1, 146) = 5.25^*$
Simulation-Based Behavior Detection Task	60.81	15.86	56.70	17.29	$F(1, 153) = 2.36$
Video-Based Passenger Observation Task	61.23	18.03	56.03	16.57	$F(1, 188) = 3.40^*$
SVT—Simultaneous *Detection Sensitivity*	3.14	1.22	3.14	1.01	$F(1, 72) = 0.01$
SVT—Simultaneous *Decrement*	0.35	0.79	0.22	0.78	$F(1, 70) = 0.45$
SVT—Successive Condition *Detection Sensitivity*	3.47	0.96	3.50	0.80	$F(1, 66) = 0.02$
SVT—Successive *Decrement*	0.22	0.70	0.43	0.53	$F(1, 61) = 1.77$

* $p < .05$

positions may be a result of task-based cognitive interference rather than an inability sustain attention and visually detect behavioral indicators. Further results showed that there were no other statistically significant differences between BDOs and TSOs on the BD battery.

Next, in a set of exploratory analyses, we examined whether there were significant differences in performance on the battery of visual search tasks between TSOs and BDOs with less than 1 year of experience within their positions, between TSOs and BDOs with from 1 year to less than 5 years of experience in their position, and TSOs and BDOs with more than 5 years of experience in their position. The purpose of running this additional set of tests was to see if the battery could differentiate between TSOs who have differing levels of job experience searching for luggage images and between BDOs who have differing levels of job experience engaging in BD tasks.

For TSOs, results showed the simulated behavior detection task was sensitive to differences in position tenure (see **Table 12**). Specifically, results showed officers with 1 to 5 years of experience in the TSO position performed significantly better on the simulation-based

behavior detection task (M = 66.61, SD = 13.58) compared to officers with 5 to 10 years of experience (M = 56.91; SD = 16.43). Officers with less than one year of experience however, did not perform better (or worst) on the task compared to officers with more than 5 years of experience or officers with 1 to 5 years of experience.

Table 12. Means, Standard Deviations, and Results of the One-way ANOVAs Comparing Performance on BD Battery Visual Search Tasks

	0–1 Years of Experience		1–5 Years of Experience		5 or More Years of Experience		ANOVA
Performance Variable	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	***F***
Luggage Image Search Task *Accuracy*	80.91	8.30	78.22	9.78	75.63	12.33	F(2, 78) = 1.16
Luggage Image Search Task *Detection Sensitivity*	1.95	.73	1.91	.67	1.67	.82	F(2, 78) = 1.02
Simulation-Based Behavior Detection Task	61.97	15.43	66.61[a]	13.58	56.91[a]	16.43	F(2, 85) = 3.48*
Video-Based Passenger Observation Task	68.17[a]	16.25	64.84[b]	16.24	56.63[ab]	18.80	F(2, 106) = 3.86*
SVT—Simultaneous *Accuracy*	0.94	0.05	0.90	0.18	0.89	0.19	F(2, 106) = 0.21
SVT—Simultaneous *Detection Sensitivity*	3.63	.59	3.21	1.23	2.93	1.34	F(2, 39) = 0.70
SVT—Simultaneous *Response Bias*	0.22	.018	0.15	0.35	.025	.039	F(2, 39) = 0.34
SVT —Simultaneous *Decrement*	0.29	1.11	0.43	0.70	0.27	0.83	F(2, 38) = 0.19
SVT—Successive *Accuracy*	0.98	0.03	0.99	0.01	0.93	0.07	F(2, 33) = 2.61
SVT—Successive *Detection Sensitivity*	3.71	0.71	4.04	.32	3.22	1.08	F(2, 34) = 2.34
SVT—Successive *Response Bias*	0.28	0.39	0.26	0.11	0.31	0.31	F(2, 34) = .07
SVT—Successive *Decrement*	0.24	1.04	0.17	0.35	0.23	0.70	F(2, 34) = 0.02

* $p < .05$

Further results showed that the passenger observation task was sensitive to differences in tenure for TSOs as well. Specifically, results showed that officers with less than one year of experience (M = 68.17; SD = 16.25) and officers with 1 to 5 years of experience (M = 64.84; SD = 16.24) scored significantly higher on the task compared to officers with more than 5 years of experience (M = 56.63; SD= 18.80). Thus, while the task was sensitive to differences in tenure, the results suggest that more experienced officers were not as proficient in the task compared to less tenured officers.

For BDOs, results failed to show any performance differences between tenure levels on the BD focused tasks (see **Table 13**). However, for the TSO oriented luggage image task, results showed that officers with less than 1 year of tenure in the BDO position (M = 1.87; SD = 0.96), and officers with 1 to 5 years of experience (M = 1.73; SD = 0.76) as a BDO performed better on the luggage image task than officers with more than 5 years of experience (M = 1.25; SD = 0.75). These results suggest that officers with more recent experience as a TSO performed better on the visual search task than officers who have been long removed from the TSO position.

Table 13. Means, Standard Deviations, and Results of the One-way ANOVAs Comparing Performance on BD Battery Visual Search Tasks

	0–1 Years of Experience		1–5 Years of Experience		5 or More Years of Experience		ANOVA
Performance Variable	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	***F***
Luggage Image Search Task *Accuracy*	76.67	16.88	75.95	12.25	70.62	12.50	F(2, 62) = 1.45
Luggage Image Search Task *Detection Sensitivity*	1.87[a]	.96	1.73[b]	.76	1.25[ab]	.75	F(2, 62) = 3.69*
Simulation-Based Behavior Detection Task	60.75	15.92	62.62	12.40	52.36	18.75	F(2, 62) = 2.56
Video-Based Passenger Observation Task	60.41	8.86	60.46	14.98	52.73	19.85	F(2, 76) = 2.04
SVT—Simultaneous *Accuracy*	0.97	0.021	0.97	0.03	0.89	0.018	F(2, 26) = 1.13
SVT—Simultaneous *Detection Sensitivity*	3.61	0.49	3.48	0.49	2.92	1.07	F(2, 28) = 1.92
SVT—Simultaneous *Response Bias*	0.28	0.14	0.39	0.18	0.18	0.47	F(2, 28 = 0.99
SVT —Simultaneous *Decrement*	0.44	0.50	0.50	0.95	-0.03	0.75	F(2, 27) = 1.58
SVT—Successive *Accuracy*	0.98	0.03	0.98	0.01	0.93	0.08	F(2, 24) = 1.80
SVT—Successive *Detection Sensitivity*	3.89	0.47	3.78	0.47	3.25	0.90	F(2, 27) = 2.03
SVT—Successive *Response Bias*	0.12	0.08	0.35	0.27	0.30	0.37	F(2, 27) = 0.70
SVT—Successive *Decrement*	.66	.78	.50	.28	.37	.57	F(2, 23) = 0.49

* p < .05

Examining Predictors of Performance on the BD Battery Visual Search Tasks

Next, we explored which set of factors contributed the most toward predicting successful visual search performance on the BD battery for TSOs and BDOs. In particular, we were interested in determining whether individual characteristics measured in the battery and vigilance as measured by the SVT predicted unique variance in performance on the visual search tasks. For this analysis the following predictors were used: working memory, self-regulation, occupational self-efficacy, TSA mission commitment, boredom proneness, job boredom, stress coping strategy (all three CITS scales), need for cognition, and SVT – Successive *Detection Sensitivity*.

Previous research suggests that general intelligence, working memory, and other cognitive-oriented individual characteristics may offer better value in predicting performance on sustained attention tasks than personality variables because they assess proximal measures of attentional allocation, which are critical for maintaining performance on vigilance type tasks (Peltier & Becker, 2017; Matthews et al., 2014). Evidence also suggests that an individual's stress coping strategy (i.e., how he or she responds to the stressful demands of maintaining high levels of performance over a sustained attention event) may help explain differences in performance on vigilance tasks (Matthews et al., 2014). Likewise, there is also evidence to suggest that individuals high in boredom proneness are more susceptible to the vigilance decrement compared with individuals who are not as prone to boredom (Sawin & Scerbo, 1995).

To investigate the predictive value of individual characteristics and vigilance on visual search performance we computed three stepwise linear regressions for both positions separately. The regressions were run separately for both positions to allow for identifying a unique set of predictors for each position rather than an omnibus set of predictors of BD battery visual search skill. The first stepwise regression explored which factors contributed the most toward predicting detection sensitivity on the luggage image search task. The second stepwise regression explored which factors contributed the most toward predicting successful performance on the video-based passenger observation task. The third stepwise regression explored predicting success on the simulation-based behavior detection task. From a practical standpoint, this analysis allowed us to determine how impactful each variable was toward explaining variance in visual search performance.

Factors Impacting BD Battery Visual Search Performance for TSOs

A stepwise regression was used due to the exploratory nature of this analysis. Stepwise regression analyses, which is a combination of forward and backward variable selection, allows for the predictive contribution of each variable to be tested each time a new variable is added to the model (Pedhazur, 1997). ***Table 14*** contains the results of the stepwise regressions for TSOs. The table includes several pieces of information. The first column shows the dependent variable in the blue row and the significant predictors in the following

rows. The second column shows the unstandardized regression coefficients (β) for each significant independent variable. The absolute value of β indicates the strength of the relationship between the variable and performance on the dependent variable. The third column contains the results of the significance test for the significant variable. A significant t-value indicates that the variable was significantly related to dependent variable performance. The fourth column shows the overall amount of variance (R^2) explained in the dependent variables by the predictor variable(s). That is, it shows how strongly the variable(s) impacted performance on the dependent variable.

Table 14. Results of Stepwise Regression Analysis for TSOs

Variable	β	t	R^2
Luggage Image Search Task			0.13
OSPAN—Working Memory	0.01	2.19*	
Video-based Passenger Observation Task			0.17
SVT—Successive *Detection Sensitivity*	0.43	2.55*	
Simulation-based Behavior Detection Task			0.25
SVT—Successive *Detection Sensitivity*	8.85	3.30*	

* $p < 0.05$; Beta values are unstandardized

The results of the stepwise regression for TSOs showed that working memory was the only factor that contributed significantly to performance on the luggage image search task, F (1,32) = 4.79, p = 0.04 and accounted for 13 percent of the variance performance. Specifically, for every one unit increase in working memory, luggage image *detection sensitivity* improved by 0.01 points. Or put another way, for every standard deviation change in working memory (SD = 17.69) there was a 0.18 change in luggage image *detection sensitivity.* Given that luggage image *detection sensitivity* scores range from 0.25 to 3.68, these changes are relatively small. An explanation of this finding could be that TSOs who were better able to hold in mind what threats he or she had just been trained to search for performed better on this task.

SVT—Successive *Detection Sensitivity* emerged as the sole significant contributor to performance on the video-based passenger observation task, F(1, 32) = 6.50, p = 0.16, explaining 17 percent of the variance in performance, and on the simulation-based passenger observation task, F(1, 32) = 10.87, p < 01, explaining 25 percent of performance variance. For every standard deviation change in SVT—Successive *Detection Sensitivity* (SD = 1.22) there was 0.52 increase in video-based passenger observation scores (range: 10 - 100) and a relatively larger 10.80 increase in simulation-based behavior detection scores (range: 19.42 - 88.89). This suggests that TSOs who were able to remain intensively and selectively focused on these two tasks performed better, especially on the

simulation-based behavior detection task. This is a logical finding given that the two tasks require intense concentration and uninterrupted visual contact with the task so as not to miss any behavioral or appearance indicators.

Factors Impacting BD Battery Visual Search Performance for BDOs

Another set of three stepwise regressions were conducted for BDOs using the set of factors identified above. The regression statistics are presented in ***Table 15***. The results showed that for BDOs, SVT—Successive *Detection Sensitivity* was the sole significant contributor to the luggage image search performance, $F(1, 26) = 5.58$, $p = 0.03$, and simulation-based behavior detection, $F(1, 25) = 8.31$, $p = .01$, accounting for 18 percent and 25 percent of variation in performance, respectively. Substantively, as with TSOs, each standard deviation change in SVT—Successive *Detection Sensitivity* (SD = 1.01) resulted in a relatively small 0.41 change in luggage image *detection sensitivity* (range: -0.14 – 3.39). Unlike TSOs, who needed higher working memory to perform well on luggage image search, it seems BDOs who were vigilant in their attention to the luggage image search task performed better. This could be because while TSOs simply had to remember what they were searching for, BDOs, who may not have searched luggage images as recently as TSOs, had to concentrate intently on the task to do well.

A one standard deviation change in SVT—Successive *Detection Sensitivity* score meant a 9.16 increase in simulation-based behavior detection score (range: 18.75 – 91.09). As with TSOs, this suggests that SVT—Successive *Detection Sensitivity* is a fairly robust predictor of this task. This indicates that for both positions the simulation-based behavior detection task may not have been sensitive to differences in BD ability but instead was sensitive to an officer’s ability to maintain uninterrupted attention on a task.

Table 15. Results of Stepwise Regression Analysis for BDOs

Variable	β	t	R^2
Luggage Image Search Task			0.18
SVT—Successive *Detection Sensitivity*	0.41	2.36*	
Video-based Passenger Observation Task			0.59
Need for Cognition	0.91	3.14*	
Occupational Self-Efficacy	-2.00	-3.23*	
O-SPAN—Working Memory	0.46	2.77*	
Simulation-based Behavior Detection Task			0.25
SVT—Successive *Detection Sensitivity*	9.07	2.89*	

* $p < 0.05$; Beta values are unstandardized

Working memory did emerge as a significant predictor of BDO performance on the video-based passenger observation task as did need for cognition and occupational self-efficacy, $F(3, 24) = 11.54$, $p < .001$. Together these three factors accounted for 59 percent of the variance in video-based passenger observation performance. Interestingly, occupational self-efficacy was related to performance in the negative direction. Meaning, officers with lower occupational self-efficacy did better on the video-based passenger observation task. Being that this relationship did not appear elsewhere, this could simply be a spurious relationship. On the other hand, it could be that those officers who do not think of themselves as capable as a BDO committed additional mental resources to the task to perform well. In terms of substantive magnitude, a one standard deviation change in need for cognition (SD = 10.28) equated to a 9.35 change in video-based passenger observation score (range: 9.09 – 94.74); while a one standard deviation change in occupational self-efficacy (SD = 5.30) resulted in a 3.30 decrease in passenger observation score.

Considered together, these regression analyses indicate that those officers, both TSOs and BDOs, who were best able to remain fully attentive to the BD battery visual search tasks had the highest performance scores. Given that the visual search battery tasks required participants to remain engaged so as not to miss a threat within a luggage image or miss a threat displayed by a passenger, just as TSOs and BDOs must do during their job activities, it is logical that vigilance emerged as a strong predictor of battery visual search skill. Working memory also emerged as a significant predictor of performance which suggests individual differences in cognitive abilities may play a significant role in task performance. Worth noting is that tasks that fall within a position's regular duty cycle (i.e., luggage image search for TSOs and passenger observation for BDOs) required less sustained attention as evidenced by vigilance not emerging as a predictor of luggage image search for TSOs nor of passenger observation for BDOs. This is a logical finding given that relative expertise on a task should result in less attentional resources needed to complete that task.

5.1.3 Research Question 3: Is performance on the Behavior Detection Battery predictive of OTJ BD performance?

Our final research question asked about the predictive nature of the BD battery as it relates to OTJ BD performance of BDOs. To answer this question, we calculated bivariate Pearson correlations between BDO BD battery visual search task performance and ratings on the internally developed BD performance rating scale—JPRS. As noted above, the TSA-developed and currently used PEC measure of BD performance did not have enough score variation between BDOs to inform this analysis and was therefore excluded.

To conduct a nuanced analysis of the predictive nature of BD job performance by the BD battery, we calculated a mean domain rating for each of the four BD domains measured by the JPRS: specific visual search skills, general visual search skills, BD and analysis, and general skills and abilities. We also calculated a global rating mean from the two global BD

skill items. To control for any rating bias that may have been introduced during global rating, we calculated a global rating by summing the ratings across the four BD domains and then dividing by the number of items on the JPRS. For example, if a supervisor rated a BDO a 3 on all items but gave a global rating of 4, it would suggest a global rating bias. The calculated global rating would instead give the BDO rated as 3 on all items a 3 as his or her calculated global rating, thereby eliminating any global rating bias. The calculated global ratings correlated very highly, although not perfectly, with the supervisor-administered global ratings, $r(35) = .882$, $p < .001$.

As shown in **Table 16**, the SVT tasks were the best predictor of OTJ BD performance. Specifically, SVT—Successive *Accuracy* and *Detection Sensitivity* correlated significantly, with each domain mean rating score and both global rating measures. This indicates that those officers who can remain vigilant in watching for critical signals (e.g., threat behaviors) while ignoring noncritical, distracting signals are more likely to be higher performing BDOs as rated by the JPRS.

Table 16. Significant Correlations between Visual Search Tasks and OTJ BD Performance

	Specific Visual Search Skills	General Visual Search Skills	Behavior Detection and Analysis	General Skills and Abilities	Global Rating	Calculated Global Rating
Simulation-based behavior detection task	ns	ns	ns	.30*	ns	ns
SVT—Simultaneous *Detection Sensitivity*	ns	ns	ns	.45*	ns	ns
SVT—Simultaneous *Decrement*	ns	-.66**	ns	ns	-.45*	-.65*
SVT—Successive *Accuracy*	.63**	.68**	.77**	.68**	.59**	.73**
SVT—Successive *Detection Sensitivity*	.42*	.53*	.67**	.56**	.48*	.64**

* p< .05; ** p < .01; ns = not statistically significant

Interestingly, the simulation-based BD task, which was designed to be the truest BD work sample test included in the BD battery, only correlated with general skills and abilities. This could mean several things. The BD skills rated within the general skills and abilities section of the JPRS are not visual search related skills. Behavioral statements in that section include, "Accomplishes tasks in a thorough and precise manner; double checks the accuracy of information and work products", "Completes written reports in a satisfactory manner", "Demonstrates sufficient effort on most tasks and assignments". One explanation is that this JPRS section is measuring an underlying trait, perhaps attention to detail, that was also

needed to perform well on the simulation-based behavior detection task. Giving credence to this explanation is the significant correlations between ratings on general skills and abilities and the SVT tasks, which require focused, sustained attention.

Another explanation for the lack of significant correlations with the simulation-based behavior detection task is that the task may not have been sensitive to differences in officer BD visual search skill levels. This could be due to using behavioral indicators that are not what BD certified officers are trained to look for, passengers in the simulated queue displaying too many indicators, or indicators being too obvious, among other things.

Next, we computed bivariate Pearson correlations between individual characteristics and JPRS ratings. As noted in **Table 17**, need for cognition was the strongest predictor of OTJ BD performance. It correlated significantly with each domain mean rating score and both global rating measures. This indicates that officers who excel at BD have a high need for cognition relative to those who do not perform as well on BD. Relatedly, those who are less prone to boredom tend to perform better as a BDO. SSSQ—Distress also correlated significantly with OTJ BD performance. This suggests that officers who are not distressed by stressful events (i.e., have lower distress scores) have higher OTJ BD performance scores. The final individual characteristic that correlated well with OTJ BD performance was CITS—Avoidance-Focused in the negative direction as would be expected. Meaning that officers who do not cope with stressful events by avoiding the stress (i.e., low CITS—Avoidance-Focused scores) tended to have higher OTJ BD performance ratings.

Table 17. Significant Correlations between Individual Characteristics and OTJ BD Performance

	Specific Visual Search Skills	General Visual Search Skills	Behavior Detection and Analysis	General Skills and Abilities	Global Rating	Calculated Global Rating
Attentional Self-Regulation	$.27^{*}$	ns	ns	ns	ns	ns
Boredom Proneness	ns	$-.33^{*}$	ns	$-.31^{**}$	$-.27^{*}$	ns
Need for Cognition	$.27^{*}$	$.41^{**}$	$.41^{**}$	$.39^{**}$	$.38^{**}$	$.41^{**}$
SSSQ—Distress	ns	ns	ns	$-.34^{**}$	$-.32^{*}$	ns
SSSQ—Engagement	ns	ns	ns	$.28^{*}$	ns	ns
CITS—Avoidance-Focused	$-.30^{*}$	ns	$-.30^{*}$	$-.40^{**}$	$-.27^{*}$	ns

* $p < .05$; ** $p < .01$; ns = not statistically significant

Finally, we computed bivariate Pearson correlations between demographics gathered in the BD battery and JPRS ratings. Pay grade and frequency of examining passengers as part of

BD were the only demographic items that significantly correlated with OTJ BD performance (see **Table 18**). These are logical findings. One would expect those who are better at BD to be promoted to higher pay grades, and that those officers who conduct BD more often to be better at BD.

Table 18. Significant Correlations between Demographics and OTJ BD Performance

	Specific Visual Search Skills	General Visual Search Skills	Behavior Detection and Analysis	General Skills and Abilities	Global Rating	Calculated Global Rating
Pay Grade	.36**	ns	.27*	.33*	.31**	ns
Frequency of examining passengers as part of BD	.25*	ns	ns	ns	.29*	ns

* p < .05; ** p < .01; ns = not statistically significant

Exploratory Analysis to Identify Predictors of JPRS Global Ratings

To further investigate the predictive value of the BD battery on OTJ BD performance, we conducted a stepwise linear regression to determine the amount of variance in OTJ BD performance explained by the significant BD battery visual search task and individual characteristic correlates of OTJ BD performance. Supervisor-assessed global rating was used as the dependent variable in the analysis because it is most like the type of rating to be used in a real-world scenario (i.e., a supervisor rating). Based on the outcome of the correlation analyses, we selected the 7 variables that were significantly correlated with OTJ BD performance to be included in the stepwise linear regression. These variables were SVT—Simultaneous *Decrement*, SVT—Successive *Accuracy*, SVT—Successive *Detection Sensitivity*, boredom proneness, need for cognition, SSSQ—Distress, and CITS—Avoidance-Focused. Since this study is focused on the predictive nature of the assessments in the developed BD battery, significant demographic correlates were not included in the regression model.

Prior to conducting the regression analysis, relevant regression assumptions were tested. Predictors were checked for multicollinearity by investigating variance inflation factors (VIF) to ensure none of the predictors were highly correlated with each other. One of the variables, SVT—Successive *Detection Sensitivity*, approached the cut off criterion of 10.0 (VIF = 9.28), which indicates that multicollinearity was an issue for that variable. This makes sense since SVT detection sensitivity is another way to measure SVT accuracy. Further investigation revealed that the correlation between SVT—Successive *Detection Sensitivity* and SVT—Successive *Accuracy* was nearly perfect ($r(61) = .97$, $p < .001$). SVT—

Successive *Detection Sensitivity* was therefore dropped from further consideration. The dataset was screened for extreme outliers, but none were identified.

Regression statistics are presented in **Table 19**. The results showed that SVT—Successive *Accuracy* was the only variable to emerge as a predictor of global ratings, F (1, 21) = 11.14, $p = 0.003$ and accounted for 35 percent of variation in OTJ BD global ratings. This suggests that officers who can sustain attention and intently focus on watching for threats over an extended period, just as BDOs must do on the job, tend to better at BD.

Table 19. Results of the Stepwise Regression Analysis of OTJ BD Performance—Global Rating Mean

Variable	β	t	R^2
Step 1			0.35
SVT—Successive *Accuracy*	7.73	3.34*	

* $p < 0.05$; Beta values are unstandardized

Exploratory Analysis to Identify Predictors of JPRS Calculated Global Ratings

Because the visual search tasks within the BD battery correlated slightly higher with calculated global ratings from the JPRS, and because there was a small amount of variance between the supervisor-applied global rating and the calculated global rating, we also conducted a stepwise regression using JPRS Calculated Global Rating as the dependent variable. Since only one individual characteristic correlated significantly with JPRS calculated global ratings, need for cognition, only four variables were used in the stepwise regression for calculated global ratings. These were SVT—Simultaneous *Decrement*, SVT—Successive *Accuracy*, SVT—Successive *Detection Sensitivity*, and need for cognition.

Results are shown in **Table 20**. As with the regression analysis using the supervisor applied global rating, SVT—Successive *Accuracy* was the sole variable to emerge as a predictor of calculated global ratings, F (1, 16) = 5.6, $p < 0.01$. SVT—Successive *Accuracy* accounted for 53 percent of variation in OTJ BD calculated global ratings. Here again, this suggests that being able to selectively and keenly watch for threats is associated with better performance on BD.

Table 20. Results of the Stepwise Regression Analysis of OTJ BD Performance—Calculated Global Rating Mean

Variable	β	t	R^2
Step 1			0.53
SVT—Successive *Accuracy*	9.20	4.23**	

** $p < 0.01$; Beta values are unstandardized

6. DISCUSSION

The purpose of this study was to develop a battery of threat detection tasks tailored to the BDO position and (1) examine whether individual differences in cognitive and non-cognitive characteristics, as measured through individual characteristics surveys, linked to performance on the BD battery; (2) determine whether there was an association between performance on the luggage screening and BD visual search tasks and identify which individual characteristics predict performance on the BD tasks; and (3) determine whether performance on the BD battery was predictive of supervisor ratings of task performance. The BD battery contained a simulation-based BD task that served as a work sample test of the threat detection and sustained attention skills employed in BD, a video-based passenger observation task that replicated the BD tasks performed during BD and assessed visual attention and passenger observation skills, a luggage screening task that assessed the visual search skills of officers, and an SVT that assessed officers' sustained attention skills.

Regarding the first research question, results showed that several individual characteristics were related to BDOs' and TSOs' performance on the BD battery. Notably, for TSOs, higher working memory capacity, which was assessed using the OSPAN, was associated with more accurate performance on the video-based passenger observation task and the SVT. For BDOs, higher working memory capacity was linked to higher accuracy scores on the luggage image visual search task. It is interesting to highlight that working memory capacity played a significant role in how well TSOs performed on the BD-oriented task and how well BDOs performed on the TSO-oriented task. Several BDOs commented that it had been years since they looked at a luggage image. Given their lack of recent experience searching luggage images for concealed threats, it seems reasonable to expect these officers to perform more poorly on the task than TSOs. Results of the current study suggest that working memory capacity may have buffered any expected performance decrements. For BDOs, having greater working memory skills may have allowed officers to quickly and accurately memorize the set of target objects required in the luggage screening task and thus perform well on the screening task. Similarly, for TSOs, having a higher working memory capacity may have allowed these officers to quickly learn and remember the set of target behaviors used in the simulation-based BD task, which may have helped them perform better compared to TSOs with lower working memory capacity. This would suggest that a working memory task might serve as a useful tool for training cross-functional skills.

Results also showed that BDOs with higher self-reported levels of need for cognition performed better on the visual search task and SVT—Successive condition than BDOs with lower levels of need for cognition. In addition, higher levels of boredom proneness among BDOs was negatively related to performance on the simulation-based BD task. These results suggest that BDOs who were more boredom prone did not perform optimally on the task. Previous research has shown a similar linkage between boredom proneness and

performance on similar types of sustained attention task. Specifically, Sawin and Scerbo (1995) found that low-boredom proneness subjects outperformed high-boredom process subjects on a 30-minute vigilance task and reported less state boredom induced by the task. Results from our study suggest that a similar relationship exists between boredom proneness and performance on the simulation-based BD task for BDOs. Notably, need for cognition and boredom proneness were significantly negatively correlated for BDOs ($r(84) = -.27$, $p < .05$). This negative correlation suggests that BDOs with higher levels of need for cognition are less likely to become boredom prone when performing a sustained attention task.

Another interesting outcome of this study was that among BDOs, higher attentional self-regulation skills were related to higher accuracy scores on the successive SVT. Specifically, results showed that BDOs with higher self-reported levels of attentional regulation skills performed better on the SVT compared to BDOs with lower levels of attention regulation skills. The SVT is a very demanding psychophysiological task and despite its short length research has shown that performance on the SVT is related to performance on much longer vigilance tasks (Shaw et al., 2010; Temple et al., 2000). Attentional regulation did not emerge as a predictor of performance for any of the other tasks within the BD battery. Perhaps the reason for this lack of transfer is that the other tasks emphasized threat detection performance and were not as psychophysically demanding as the SVT, which required participants to engage in a sustained attention state long enough to induce a vigilance decrement.

Further results showed that task-focused coping as measured by the CITS—Task-Focused scale provided additional insights into how well BDOs performed on the simulation-based BD task. Specifically, results showed BDOs who engaged in more task-focused coping behaviors, such as putting more effort toward the task, tended to perform better on the simulation-based BD task than individuals who disengaged from the task or did not put forth as much effort toward the task. Items that assessed task-focused coping included "Worked out a strategy for successful performance," "Concentrated hard on doing well," "Made every effort to achieve my goals," and "Was determine in my efforts to overcome any problems." Task-focused coping also emerged as a significant predictor of performance for TSOs on the visual search task and SVT. It is interesting that task-focused coping was related to performance for BDOs and TSOs on the work sample tasks that simulated the tasks and duties they are familiar with performing. This suggests that individuals high in task-focused coping may have used a specific strategy while performing the task or simply put forth more effort toward performing well on the task compared to individuals who engaged in more avoidance-based coping.

For TSOs, job boredom showed an unexpected significant positive correlation with accuracy on the luggage screening task. It is important to note that the Job Boredom Scale measured antecedents of boredom in the workplace and included items such as "Do you often get

bored with your work," "Is your work monotonous," and "Does your job seem repetitive?" Thus, one possible explanation for the positive correlation is that some TSOs have acquired significant expertise in luggage screening and therefore find conducting luggage search easy. The ease of the task in turn causes them to feel increased levels of job boredom. This explanation is rooted in previous research that suggests that job tenure is related to job boredom proneness and the assumption that luggage accuracy is linked to increased job tenure (Drory, 1982).

An alternative explanation for this unexpected relationship, is that the positive relationship is an artifact of the scoring approach used to measure visual search accuracy on the luggage screening task. As previously noted, accuracy was calculated based on the total number of trials in which a participant correctly identified an image with an IED present (i.e., hit), and the number of trials in which he or she correctly dismissed an image that did not contain an IED (i.e., correct rejection). One limitation with this scoring approach is that it rewards participants who adopt a more conservative response strategy. That is, participants who respond less often are at an advantage because they accumulate more correct rejection points than participants who adopt a more liberal decision-making threshold.

Indeed, results of an exploratory set of correlational analyses showed that correct rejection was significantly positively correlated with accuracy scores ($r = 0.96$) and false alarms were negatively correlated with accuracy scores ($r = -0.96$). The relationship between hits and accuracy was positive but much weaker ($r = 0.34$). This suggests that accuracy was highly influenced by the number of correct rejections an individual had and negatively impacted by false alarms. Correlation analyses also showed that TSOs with higher self-reported levels of job boredom were less likely to commit false alarms and had higher levels of correct rejections, which suggests that these officers may have adopted a more conservative response threshold. Thus, it appears that job boredom does not indicate that an officer is good at identifying targets, but rather suggests that an individual is more likely to adapt a conservative response threshold, which in the case of this study led to higher visual search accuracy scores due to manner in which accuracy was calculated. Notably, there was a significant correlation between job boredom and response bias on the SVT, but there was not a significant relationship between boredom proneness and response.

For TSOs, results also showed that stress coping was significantly related to performance on several items from the battery. Specifically, there were significant positive correlations between task-focused coping and luggage image visual search task accuracy and detection sensitivity. Conversely, results showed negative correlations between avoidance and emotion-focused coping for TSOs and performance on the video-based passenger observation task and SVT. Specifically, TSOs with higher reported levels of emotion-focused coping had lower scores on the video-based passenger observation task, and TSOs who engaged in more avoidance-focused coping had lower scores on the SVT. Thus, the CITS

measure provided useful insights for predicting performance on other visual search and sustained attention tasks. Our results suggest that individuals who engage in more task-focused coping strategies are more likely to perform well on other types of sustained attention tasks, whereas individuals who engage in avoidance or emotion-focused coping are less likely to perform well on BD tasks.

Items from the SSSQ failed to correlate with any of the tasks in the battery. Although previous research has shown that stress states can predict performance on vigilance and sustained attention tasks, perhaps the tasks used in our study were not performed long enough to allow the measure to detect differences in performance between the tasks or between BDOs and TSOs.

Regarding the second research question, results showed that performance on the BD and luggage image visual search tasks were positively correlated with each other. Specifically, for TSOs, high performance on the luggage screening task was associated with more accurate performance on the simulation-based BD task, more accurate performance on the video-based passenger observation task, and more accurate performance on the simultaneous and successive versions of the SVT. The same pattern of results also emerged for BDOs. Specifically, officers who performed well on the simulation-based BD and video-based passenger observation tasks also performed well on the luggage image screening and SVT tasks. These results suggest that for both sets of officers, being good at one task was associated with being good at the others. Regarding the task battery, these results suggest that the tasks, although distinct, measured similar psychological constructs.

Additional analyses that examined differences in performance between BDOs and TSOs on the battery showed that, in general, TSOs scored higher on the luggage screening visual search and video-based passenger observation task compared to BDOs. TSOs also scored higher compared to BDOs on the simulation-based BD simulation, but the difference was not statistically significant. One hypothesis as to why TSOs scored higher on the video-based passenger observation task compared to BDOs is that BDOs experienced proactive interference, and this interference led to lower scores on the test. Indeed, many BDOs commented that the passenger observation tasks were difficult to perform because they had to quickly "block out" the cues they had been certified to look for and then quickly learn a new set of cues to perform the task. Perhaps this interference led to the observed differences in performance. TSOs, on the other hand, did not report experiencing any form of interference and were able to engage in the task without needing to override an existing performance schema. Further analyses showed that there were no differences in performance between BDOs and TSOs on the simultaneous or successive versions of the SVT.

Results of regression analyses examining which individual characteristics explained the most variance in performance on the BD battery visual search tasks showed that for TSOs,

working memory capacity was the best indicator of performance on the luggage image search task. Given that participants were instructed to search for a limited set of battery operated IEDs and not a full array of potential threats like would be encounter in their normal work environment, this finding likely suggests that those TSOs who best remembered the IED threats specifically called out in the task were able to performance at a higher level. Accuracy performance on the SVT—Successive condition was the best indicator of TSO performance on the battery's BD tasks. Given the importance of sustained and selective attention skills in performing BD tasks and similar watch-keeping tasks, this finding is not surprising.

Regression analyses for BDOs found that vigilance performance on the SVT—Successive task was the best predictor of performance on the luggage image search task and the simulation-based BD task. These results suggest that vigilance performance is a strong predictor of performance on threat detection and BD tasks and may offer predictive value. Need for cognition, occupational self-efficacy, and working memory all emerged as significant predictors of video-based passenger observation task, accounting for over half the variance in performance. This finding is unique in the context of this study in that this the only time occupational self-efficacy shows as being related to performance on a visual search task. A possible explanation is that those officers who are not confident in their abilities but who enjoy mentally challenging tasks were able to commit cognitive resources to succeeding on the task, whereas officers who are confident in their abilities may have been over confident when completing this task and therefore did not commit sufficient mental resources.

Regarding the final research question, results showed that performance on the SVT task, specifically the successive condition, was the best predictor of OTJ BD performance. SVT—Successive showed significant correlations with all measures of OTJ BD performance as measured by the in-house developed JPRS. These measures included specific search skills, general search skills, BD and analysis skills, global skill ratings, and a calculated global skill rating. Likewise, the individual characteristic need for cognition had significant positive correlations with each of the OTJ measures. Taken together, along with the finding that boredom proneness was significantly negatively correlated with three of the five OTJ measures, these findings indicate that officers with a need for cognitively demanding activities who can remain engaged in that activity (i.e., are vigilant and do not become bored) are more likely to be successful at BD. This is a logical finding given the psychologically demanding and attention demanding nature of BD. **Table 21** provides a reference for how these performance correlates relate to the KSAs for the BDO position.

Table 21. Linkage between Predictive Behavior Detection Battery Tasks and KSAs

	Attention to Detail	Situation Awareness	Sustained Attention	Selective Attention	Memory
SVT	X		X	X	X
Need for Cognition	X		X	X	X
Boredom Proneness (Lack of)	X	X	X	X	

Regression analyses showed that performance on the SVT—Successive condition accuracy was the best predictor of OTJ BD performance regardless of whether the supervisor assess global BD skill rating or a calculated global BD skill rating was used as the measure of OTJ BD performance. SVT—Successive accuracy explained 35 to 53 percent of the variance in OTJ BD performance depending on which global rating metric was used. This finding suggests that adding a vigilance measurement task may improve the predictive validity of a BD selection or training program.

6.1 Implications for TSA and Directions for Future Research

The BD task battery used in the present study included a modified version of a video-based passenger observation task, a newly developed simulation-based BD task, and a luggage image screening task. It also included a shorted vigilance task that has been frequently used in vigilance research (see Matthews et al., 2014; Shaw et al., 2010; Temple et al., 2000). Results from the current study suggest that the tests within the BD battery represent unique but overlapping constructs. Scores on the two BD-oriented tasks, the vigilance task, and the luggage screening tests were positively correlated with one another. Thus, in response to the research question "Are individuals who are good at one task also good at the other?" our results suggest that the answer is yes. The BD battery showed promising potential for identifying individuals who could engage in hybrid screening roles or identifying TSOs who might excel at the BD function.

In terms of the BD job function, in the near term, TSA could use the results of this study to look at non-BD certified TSOs who are good at luggage image visual search as potential candidates for BD certification. Thereafter in the medium-term, TSA could consider adding to their BD certification selection process measures of individual cognitive differences, specifically a measure of need for cognition, boredom proneness, and stressful event coping strategies, as all showed significant correlations with OTJ BD performance. Lastly, in the long-term, as it will take a more concerted effort to implement, TSA could consider developing or adapting a measure of vigilance for addition to the BD certification selection process. Caution should be given when considering measures of these traits and skills, as

this study investigated specific measures of these traits and skills and the relationships uncovered in this study may not generalize to other measures not studied.

Future research should continue to focus on developing valid and reliable skills-based assessments that can be used to measure BD skills. For instance, the simulation-based BD task could be enhanced in several meaningful ways. First, future research could include the behavioral indicators officers have been certified to recognize. As previously mentioned, some BD certified participants noted the difficulty of suspending their training in order to monitor passengers for just-learned behavioral indicators. This may have caused interference in the performance of BD certified officers, thereby decreasing the predictive ability of the simulation-based BD task.

A second area of future research could investigate the optimal length of the simulation-based BD task. For this effort we limited the task to a 12-minute period of performance based on procedures used by BDOs when they engage in BD activities. Perhaps increasing the length of the task to 25 or 30 minutes could offer better diagnostic sensitivity from an assessment standpoint because it could more accurately assess the sustained attention skills of job candidates in a realistic task.

A third area for future research is to include additional functions within the simulation-based BD battery that require officers to engage in casual conversation with suspicious passengers. This would allow the simulation to assess visual and nonvisual search-oriented skills. For instance, the simulation could require officers or job candidates to interview passengers if they display a certain behavior or a specific series of behavioral indicators. Upon selecting a passenger to interview, the simulation could present the officer with pre-scripted questions to ask the passengers. Interviews could include multiple branches or levels of questions to mimic real-world situations. The scripts and responses could be based on realistic accounts and current standards for engaging in resolution conversations. Incorporating this feature in the BD task could be particularly relevant for assessing the interpersonal skills of BDOs, which are also an important part of BD.

A fourth area of future research would be to create a virtual-reality (VR) or augmented reality (AR) version of the task that allows officers to engage in BD and passenger observation in a highly engaging and realistic virtual environment. Coupling a VR or AR-based application with multimodal sensors such as eye tracking could provide powerful game-trace data that could be used to identify behavioral indicators of successful performance among existing BD-certified officers and among job candidates.

A fifth avenue of future research is to use the simulation-based BD task as a platform for assessing the multitasking abilities of BD-certified screeners. For example, a secondary task, such as an auditory monitoring task, could be added to the simulation that would require officers to monitor a simulated radio signal and respond (either verbally or by hitting

a key on the keyboard) whenever they hear a target phrase or term. This dual-task paradigm could be used to examine the impact of cognitive load on BD performance.

Sixth, future research could be to test alternate simulated environments within the airport beyond the security line passenger queue to include other environs where BD could be expected to take place. This could include at the travel document check station, within the sterile section of the terminal, or even at a terminal gate. This environmental expansion could lend an additional level of realism and face validity to the task because it would align with the recent integration of the BD capability into the TSO duty cycle.

A seventh further area worth exploring is increasing the difficulty of the simulation-based BD task by adding target cues to passengers who are deeper in the passenger queue. This additional feature was suggested by several former and current BDOs during pilot testing but was not fully implemented for the current effort. Instead, the current version of the task limits the target behaviors to the first line of passengers. One of the reasons we limited the task to the first line of passengers was that we believed the task would be too difficult for job candidates if they were required to search for behavior cues two and three rows deep. Further empirical evidence is needed to determine whether increasing the difficulty of the task using this manipulation adds to the task's predictive value.

In addition to exploring ways to enhance and improve the BD simulation, another testing approach that could benefit from additional research is developing a large repository of passenger videos that could be used to improve skills-based passenger observation assessments. The repository could include videos that are meta-tagged and graded by experts to determine which indicators are contained in the videos and which passengers display each behavioral cue and how often. These videos could then be used to create a suite of passenger observation tests. This testing approach may offer high face validity among BD candidates because it uses video of actual passengers. Furthermore, the large library of videos could provide a good source of content, which would allow researchers to create parallel test forms or tests of different difficulty levels. Such tests may offer more sensitivity in parsing out high performers from low performers compared to the current video-based passenger observation task.

A further testing approach would be to use actors to create videos of passengers waiting in line at a mock checkpoint. Like the simulation-based approach, this method provides a better degree of experimental control over the use of actual videos of passengers because the researcher could control which behaviors passengers' display, the subtleness of the behavior, and how frequently certain behaviors occur. As a cost-saving measure, these videos could be created at the Federal Law Enforcement Training Center during BD training wherein actor-based passengers could be used to simulate behavioral cues in a mock passenger checkpoint. Furthermore, these actor-based videos could be filmed in 360° video. When combined with VR technology, these videos become completely immersive and

lifelike. This would essentially allow a participant to be implanted into a real but experimentally controlled environment.

Finally, it is important to note that the current study had a fairly low sample size, which restricts the generalizability of the results. Future research might address this issue by collecting data from a larger sample of BDOs and TSOs from additional airport locations. Doing so would allow researchers to conduct more advanced analyses (i.e., multiple regression) and provide more generalizable results. We were also not able to fully establish the convergent validity of our in-house developed BDO job performance rating scale. Although the scale has face validity and content validity, because the domains and items were based on the PEC rating form and BDO job analyses, we were not able to statistically establish that it measures BDO performance because the PEC scores we received for participant BDOs did not have enough between-officer variation. Further research is needed to concretely establish how well the JPRS measures BD performance.

7. CONCLUSIONS

The three the primary objectives of this study were to (1) investigate whether individual differences in cognitive and non-cognitive characteristics were linked to performance on the behavior detection battery; (2) determine whether there was an association between performance on the luggage screening task and BD tasks and examine differences in performance on the BD task between BDOs and TSOs; and (3) identify whether performance on the BD battery was predictive of OTJ performance.

We found that (1) individual characteristics were related to BDOs' and TSOs' performance on the BD battery. Working memory capacity played a significant role in how well TSOs performed on the BD-oriented task and how well BDOs performed on the TSO-oriented task. (2) High performance on the luggage screening task was associated with more accurate performance on the simulation-based BD task, more accurate performance on the video-based passenger observation task, and more accurate performance on the vigilance task. These results suggest that for both sets of officers, being good at one task was associated with being good at the other tasks. (3) Performance on the vigilance task was the best predictor of OTJ BD performance. Need for cognition, low boredom proneness, low post-task distress, and low avoidance stress-coping all showed a relationship with higher OTJ BD performance.

These results suggest that TSA might consider non-BD certified TSOs who are good at luggage image visual search as potential candidates for BD certification. TSA may also consider adding to their BD certification selection process measures of individual cognitive differences, specifically a measure of need for cognition, boredom proneness, and stressful event coping strategies, as all showed significant correlations with OTJ BD performance. Lastly, TSA could consider developing or adapting a measure of vigilance for addition to the BD certification selection process.

Future research could focus on refining the simulation-based behavior detection task and developing additional exercises and tests that assess BD capabilities. By creating more sensitive tests and scoring approaches, TSA may be able to more readily identify individual differences that predict BD ability. Conducting additional basic research in this area may be particularly important as TSA develops its subset of fully BD-certified officers who engage in cross-functional tasks including luggage screening and BD.

References

Biggs, A. T., Cain, M. S., Clark, K., Darling, E. F., & Mitroff, S. R. (2013). Assessing visual search performance differences between Transportation Security Administration Officers and nonprofessional visual searchers. *Visual Cognition*, *21*(3), 330–352. http://dx.doi.org/10.1080/13506285.2013.790329

Cacioppo J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–131. http://dx.doi.org/10.1037/0022-3514.42.1.116

Drory, A. (1982). Individual differences in boredom proneness and task effectiveness at work. *Personnel Psychology*, *35*(1), 141–151. http://dx.doi.org/10.1111/j.1744-6570.1982.tb02190.x

Farmer, R., & Sundberg, N. D. (1986). Boredom proneness—the development and correlates of a new scale. *Journal of Personality Assessment*, *50*(1), 4–17. http://dx.doi.org/10.1207/s15327752jpa5001_2

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84–96.

Grau, R., Salanova, M., & Peiró, J. M. (2001). Moderating effects of self-efficacy on occupational stress. *Spanish Journal of Psychology*, 5, 63–74.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Huntington, NY: Robert E. Krieger Publishing Co.

Helton, W. S. (2004). Validation of a short stress state questionnaire. *Proceedings of the Human Factors and Ergonomics Society*, 48, 1238–1242.

Hendrickson, C., Myers, T. L., Loignon, A. C., Gilbert, S. N., Kurtessis, J., Clayton, T P., Norris, D. G., & Davies, S. A. (2012). *Updating the Personnel Selection System for Behavior Detection Officers. Validation Study Volume I: Technical Report*. Prepared for U.S. Naval Research Laboratory, Washington, DC.

International Personality Item Pool: A Scientific Collaboratory for the Development of Advanced Measures of Personality Traits and Other Individual Differences. (n.d.). Retrieved from https://ipip.ori.org/newNEOFacetsKey.htm.

Jerusalem, M., & Schwarzer, R. (1992). Self-efficacy as a resource factor in stress appraisal process. In R. Schwarzer (Ed.), *Self-efficacy: Thought control of action*. (pp. 194–213). Washington, DC: Hemisphere.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *26*(2), 336–358. http://dx.doi.org/10.1037/0278-7393.26.2.336

Kass, S. J., Vodanovich, S. J., & Callender, A. (2001). State-trait boredom: Relationship to absenteeism, tenure, and job satisfaction. *Journal of Business and Psychology*, *16*(2), 317–327. http://dx.doi.org/10.1023/A:1011121503118

Keppel, G., & Wickens, T. D. (2004). Effect size, power, and sample size. *Design and Analysis. A Researcher's Handbook, 4*, 159-801.

Lee, T.W. (1986). Toward the development and validation of a measure of job boredom. *Manhattan College Journal of Business, 15*, 22–28.

Luszczynska, A., Diehl, M., Gutiérrez-Doña, B., Kuusinen, P., & Schwarzer, R. (2004). Measuring one component of dispositional self-regulation: Attention control in goal pursuit. *Personality and Individual Differences*, *37*(3), 555–566. http://dx.doi.org/10.1016/j.paid.2003.09.026

Matthews, G., & Campbell, S. E. (1998, October). Task-induced stress and individual differences in coping. (SAGE Publications.) In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(11), 821–825. http://dx.doi.org/10.1177/154193129804201111

Matthews, G., Warm, J. S., Shaw, T. H., & Finomore, V. S. (2014). Predicting battlefield vigilance: A multivariate approach to assessment of attentional resources. *Ergonomics*, *57*(6), 856–875. http://dx.doi.org/10.1080/00140139.2014.899630

Pedhazur, E.J. (1997). *Multiple regression in behavioral research* (3rd ed.). Belmont, CA: Wadsworth.

Peltier, C., & Becker, M. W. (2017). Individual differences predict low prevalence visual search performance. *Cognitive research: principles and implications*, *2*(1), 5.

Rigotti, T., Schyns, B., & Mohr, G. (2008). A short version of the occupational self-efficacy scale: Structural and construct validity across five countries. *Journal of Career Assessment*, *16*(2), 238–255. http://dx.doi.org/10.1177/1069072707305763

Rose, C. L., Murphy, L. B., Byard, L., & Nikzad, K. (2002). The role of the Big Five personality factors in vigilance performance and workload. *European Journal of Personality*, *16*(3), 185–200. http://dx.doi.org/10.1002/per.451

RTI International. (2015a). *Behavior Detection Officer basic visual search comparative testing report: Phase I.* Prepared for the Department of Homeland Security.

RTI International. (2015b). *Job requirements summary report for the Behavior Detection Officer*. Research Report Prepared for the Department of Homeland Security.

RTI International. (2016). *Behavior detection visual search task analysis project: Draft visual search battery report.* Prepared for Department of Homeland Security.

RTI International. (2017a). *Behavior detection visual search task analysis project: Experimental Design Plan.* Prepared for Department of Homeland Security.

RTI International. (2017b). *Behavior detection visual search task analysis project: Summary of BDO on-the-job performance measures report.* Prepared for Department of Homeland Security.

Sawin, D. A., & Scerbo, M. W. (1995). Effects of instruction type and boredom proneness in vigilance: Implications for boredom and workload. *Human Factors*, *37*(4), 752–765. http://dx.doi.org/10.1518/001872095778995616

Shaw, T. H., Matthews, G., Warm, J. S., Finomore, V. S., Silverman, L., & Costa, P. T., Jr. (2010). Individual differences in vigilance: Personality, ability and states of stress. *Journal of Research in Personality*, *44*(3), 297–308. http://dx.doi.org/10.1016/j.jrp.2010.02.007

Spain, R., & Ladd, K. (2015, March). *BDO basic visual search project: State of the science summary report.* Prepared for Department of Homeland Security.

Spain, R. D., Ladd, K. M., & Hedge, J. W. (2016, March). *BDO basic visual search project: Phase 1 comparative testing final report*. Prepared for Department of Homeland Security.

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin*, 124, 240-261.

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

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics*. New York: Harper Collins.

Temple, J. G., Warm, J. S., Dember, W. N., Jones, K. S., LaGrange, C. M., & Matthews, G. (2000). The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. *Human Factors*, *42*(2), 183–194. http://dx.doi.org/10.1518/001872000779656480

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. http://dx.doi.org/10.3758/BF03192720

Wickens, C. D., & Hollands, J. (2000). *Engineering psychology and human performance* (3rd ed.). New York: Prentice Hall.

Appendix A:
Formulas for Calculating Signal Detection Theory Measures

Traditional SDT measures of detection sensitivity (d') and response bias (c) are commonly used to determine decision making and discrimination processes of human observers (Green & Swets, 1966). Sensitivity refers to an observer's ability to discriminate true signals from background noise, or more commonly, the ability to make correct detection decisions. Computationally, sensitivity is defined as the difference between the mean of the signal-plus-noise distribution and the mean of the noise distribution or,

$$d' = \Phi^{-1}[P(HI)] - \Phi^{-1}[P(FA)]$$

where,

d'	=	sensitivity
$\Phi^{-1}[P(HI)]$	=	z score corresponding to the point below which the area under the standard normal distribution equals the proportion of hits,
$\Phi^{-1}[P(FA)]$	=	z score corresponding to the point below which the area under the standard normal distribution equals the proportion of false alarms.

Response bias refers to the observer's general readiness to respond and is commonly notated as c, or computationally as,

$$c = -\frac{\Phi^{-1}[P(HI)] + \Phi^{-1}[P(FA)]}{2}$$

Appendix B:
Individual Characteristics Measures

Demographics

Please answer the following demographic questions.

Question #1

Response is required*

At which airport do you work?

- ◯ DFW
- ◯ DTW
- ◯ LGA
- ◯ PDX
- ◯ PHX
- ◯ PVD
- ◯ RDU
- ◯ TPA

Question #2

Response is required*

How long have you worked for the TSA (please respond with the number of years you've worked with TSA (e.g., 1 year)?

Question #3

Response is required*

What is your current position in the TSA?

- ◯ TSO
- ◯ Lead TSO
- ◯ Supervisory TSO
- ◯ BDO
- ◯ Lead BDO
- ◯ Supervisory BDO
- ◯ TSO certified in Behavior Detection

Question #4

Response is required*

What is your current pay grade?

- ○ E
- ○ F
- ○ G
- ○ H

Question #5

Response is required*

How long have you been in your current position (please respond with the number of years and then months you've been in your current position e.g., 2 years 6 months)?

Question #6

Response is required*

What is your age

Question #7

Response is required*

What is your sex?

- ○ Male
- ○ Female

Question #8

What is your race?

- ○ Black of African American
- ○ Native Hawaiian or Pacific Islander
- ○ Two or more races
- ○ White
- ○ Other
- ◉ No answer

Question #9

What is your ethnicity?

- ○ Hispanic or Latino
- ○ Not Hispanic or Latino
- ○ Other
- ◉ No answer

Question #10

Response is required*

What is the highest level of education you have completed?

- ○ Some High School
- ○ High School
- ○ Associate Degree
- ○ Some College
- ○ College
- ○ Some Masters
- ○ Masters
- ○ Doctorate

Question #11

Response is required*

Did you previously serve in the military?

- ○ Yes
- ○ No

Question #12

Did your previous Military service involve searches like those you now conduct for the TSA (e.g., searching crowds for threats)?

- ○ Yes
- ○ No
- ◉ No answer

Question #13

Response is required*

Do you have any Law Enforcement or security (e.g., Police, Highway Patrol) experience?

- ○ Yes
- ○ No

Question #14

Did your previous Law enforcement service involve searches like those you now conduct for the TSA (e.g., did you search for dangerous items or contraband?)?

- ○ Yes
- ○ No
- ◉ No answer

Question #15

Which type of Behavior Detection Protocol are you certified to use?

- ○ Standard Protocol
- ○ Optimized
- ○ Both
- ○ Neither
- ◉ No answer

Question # 16

Response is required*

Please rate how often you engage in the following activities using the following scale.

		Never	Very Rarely	Sometimes	Frequently	Very Frequently
Examining X-ray images of carry-on luggage at screening checkpoint	◉	option 1 ○	option 2 ○	option 3 ○	option 4 ○	option 5 ○
Examining images from advanced imaging technology (AIT)	◉	option 1 ○	option 2 ○	option 3 ○	option 4 ○	option 5 ○
Examining passenger behaviors as part of BD	◉	option 1 ○	option 2 ○	option 3 ○	option 4 ○	option 5 ○

Question # 17

When was your last certification in X-ray?

- ○ 2017 (currently certified)
- ○ 2016
- ○ 2015
- ○ 2014
- ○ 2013
- ○ More the 5 years ago
- ○ I've never been certified in X-ray
- ◉ No answer

Question # 18

When was your last certification in BD?

- ○ 2017 (currently certified)
- ○ 2016
- ○ 2015
- ○ 2014
- ○ 2013
- ○ More the 5 years ago
- ○ I've never been certified in BD
- ◉ No answer

Boredom Proneness Scale

Instructions: Please answer the following questions as they generally pertain to yourself using either true or false.

1. It is easy for me to concentrate on my activities
2. Frequently when I am working I find myself worrying about other things
3. Time always seems to be passing slowly
4. I often find myself at "loose ends," not knowing what to do
5. I am often trapped in situations where I have to do meaningless things
6. Having to look at someone's home movies or travel slides bores me tremendously
7. I have projects in mind all the time, things to do
8. I find it easy to entertain myself
9. Many things I have to do are repetitive and monotonous
10. It takes more stimulation to get me going than most people
11. I get a kick out of most things I do
12. I am seldom excited about my work
13. In any situation I can usually find something to do or see to keep me interested
14. Much of the time I just sit around doing nothing
15. I am good at waiting patiently
16. I often find myself with nothing to do, time on my hands
17. In situations where I have to wait, such as in a line, I get very restless
18. I often wake up with a new idea
19. It would be very hard for me to find a job that is exciting enough
20. I would like more challenging things to do in life
21. I feel that I am working below my abilities most of the time
22. Many people would say that I am a creative or imaginative person
23. I have so many interests, I don't have time to do everything
24. Among my friends, I am the one who keeps doing something the longest
25. Unless I am doing something exciting, even dangerous, I feel half-dead and dull
26. It takes a lot of change and variety to keep me really happy
27. It seems that the same things are on television or in the movies all the time; it's getting old
28. When I was young, I was often in monotonous and tiresome situations

Job Boredom Scale

The questions that follow all deal with your experience of your job as dull or exciting. Please answer the questions with respect to your own reactions to your present job. Instructions. Please use the scale that follows to answer the questions.

1 = Never	2 = Very rarely	3 = Sometimes	4 = Often	5 = Very Often	6 = Almost always	7 = Always

1. Do you often get bored with your work?
2. Is your work monotonous?
3. Would you like to change from your type of work to another from time to time (if the pay were the same)?
4. How well do you like the work you do?
5. Do you often get tired on the job?
6. Do you find the job dull?
7. Does the job go by slowly?
8. Do you become irritable on the job?
9. Do you get apathetic on the job?
10. Do you get mentally sluggish during the day?
11. Do you get drowsy on the job?
12. Does the time seem to go by slowly?
13. Are there long periods of boredom on the job?
14. Does the job seem repetitive?
15. During the day, do you think about doing another task?
16. Does monotony describe your job?
17. Is your work pretty much the same day after day?

IPIP-NEO-Conscientiousness Scale

How Accurately Can You Describe Yourself?

Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Indicate for each statement whether it is 1. Very Inaccurate, 2. Moderately Inaccurate, 3. Neither Accurate Nor Inaccurate, 4. Moderately Accurate, or 5. Very Accurate as a description of you.

		Very Inaccurate (1)	Moderately Accurate (2)	Neither Accurate nor inaccurate (3)	Moderately accurate (4)	Very accurate (5)
+ keyed	Am always prepared.					
	Pay attention to details.					
	Get chores done right away.					
	Carry out my plans.					
	Make plans and stick to them.					
	Complete tasks successfully.					
	Do things according to a plan.					
	Am exacting in my work.					
	Finish what I start.					
	Follow through with my plans.					
– keyed	Waste my time.					
	Find it difficult to get down to work.					
	Do just enough work to get by.					
	Don’t see things through.					
	Shirk my duties.					
	Mess things up.					
	Leave things unfinished.					
	Don’t put my mind on the task at hand.					
	Make a mess of things.					
	Need a push to get started.					

Need for Cognition Scale

For each of the statements below, please indicate whether or not the statement is characteristic of you or of what you believe. For example, if the statement is extremely uncharacteristic of you or of what you believe about yourself (not at all like you) please place a "1" on the line to the left of the statement. If the statement is extremely characteristic of you or of what you believe about yourself (very much like you) please place a "5" on the line to the left of the statement. You should use the following scale as you rate each of the statements below.

1 extremely uncharacteristic of me	2 somewhat uncharacteristic of me	3 uncertain	4 somewhat characteristic of me	5 extremely characteristic of me

1. I prefer complex to simple problems.
2. I like to have the responsibility of handling a situation that requires a lot of thinking.
3. Thinking is not my idea of fun.**
4. I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.**
5. I try to anticipate and avoid situations where there is a likely chance I will have to think in depth about something.**
6. I find satisfaction in deliberating hard and for long hours.
7. I only think as hard as I have to.**
8. I prefer to think about small daily projects to long term ones.**
9. I like tasks that require little thought once I've learned them.**
10. The idea of relying on thought to make my way to the top appeals to me.
11. I really enjoy a task that involves coming up with new solutions to problems.
12. Learning new ways to think doesn't excite me very much.**
13. I prefer my life to be filled with puzzles I must solve.
14. The notion of thinking abstractly is appealing to me.
15. I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.
16. I feel relief rather than satisfaction after completing a task that requires a lot of mental effort.**
17. It's enough for me that something gets the job done; I don't care how or why it works.**
18. I usually end up deliberating about issues even when they do not affect me personally.

Note: **=reverse scored item.

SSSQ

General Instructions. This questionnaire is concerned with your feelings and thoughts RIGHT NOW. We would like to build up a detailed picture of your current state of mind, so there are quite a few questions, divided into four sections. Please answer every question, even if you find it difficult. Answer, as honestly as you can, what is true of you. Please do not choose a reply just because it seems like the 'right thing to say'. Your answers will be kept entirely confidential. Also, be sure to answer according to how you feel AT THE MOMENT. Don't just put down how you usually feel. You should try and work quite quickly: there is no need to think very hard about the answers. The first answer you think of is usually the best.

For each statement, please choose one of the following answers, according to how much you agree with the statement.

Extremely = 4 Very much = 3 Somewhat = 2 A little bit = 1 Not at all = 0

I feel dissatisfied.	0 1 2 3 4
I feel alert.	0 1 2 3 4
I feel depressed.	0 1 2 3 4
I feel sad.	0 1 2 3 4
I feel active.	0 1 2 3 4
I feel impatient.	0 1 2 3 4
I feel annoyed.	0 1 2 3 4
I feel angry.	0 1 2 3 4
I feel irritated.	0 1 2 3 4
I feel grouchy.	0 1 2 3 4
I am committed to attaining my performance goals	0 1 2 3 4
I want to succeed on the task	0 1 2 3 4
I am motivated to do the task	0 1 2 3 4
I'm trying to figure myself out.	0 1 2 3 4
I'm reflecting about myself.	0 1 2 3 4
I'm daydreaming about myself.	0 1 2 3 4
I feel confident about my abilities.	0 1 2 3 4
I feel self-conscious.	0 1 2 3 4
I am worried about what other people think of me.	0 1 2 3 4
I feel concerned about the impression I am making.	0 1 2 3 4
I expect to perform proficiently on this task.	0 1 2 3 4
Generally, I feel in control of things.	0 1 2 3 4
I thought about how others have done on this task.	0 1 2 3 4
I thought about how I would feel if I were told how I performed.	0 1 2 3 4

Coping Inventory for Task Stress (CITS)

Think about how you usually react to potentially difficult, stressful or upsetting situations where you have to carry out some task that requires mental effort, concentration or intensive thought. Think especially about activities you have to do, rather than those you enjoy doing. These might include studying, taking tests and examinations, driving/cycling, working with other people and managing your financial affairs. Below are listed some options for dealing with problems that may occur in mentally demanding situations of these kinds. Please circle a number from 0 to 5 for each item, to indicate how much you TYPICALLY engage in each type of activity when you encounter a stressful situation that requires a high level of mental effort, specifically as a deliberately chosen way of dealing with problems. To answer circle one of the following answers:

Extremely = 4 Very much = 3 Somewhat = 2 A little bit = 1 Not at all = 0

I …

1.	Work out a strategy for successful performance	0 1 2 3 4
2.	Worry about what I will do next	0 1 2 3 4
3.	Stay detached or distanced from the situation	0 1 2 3 4
4.	Decide to save my efforts for something more worthwhile	0 1 2 3 4
5.	Blame myself for not doing better	0 1 2 3 4
6.	Become preoccupied with my problems	0 1 2 3 4
7.	Concentrate hard on doing well	0 1 2 3 4
8.	Focus my attention on the most important parts of the task	0 1 2 3 4
9.	Act as though the task wasn’t important	0 1 2 3 4
10.	Don’t take the task too seriously	0 1 2 3 4
11.	Wish that I could change what was happening	0 1 2 3 4
12.	Blame myself for not knowing what to do	0 1 2 3 4
13.	Worry about my inadequacies	0 1 2 3 4
14.	Make every effort to achieve my goals	0 1 2 3 4
15.	Blame myself for becoming too emotional	0 1 2 3 4
16.	Am single-minded and determined in my efforts to overcome any problems	0 1 2 3 4
17.	Give up the attempt to do well	0 1 2 3 4
18.	Tell myself it isn’t worth getting upset	0 1 2 3 4
19.	Am careful to avoid mistakes	0 1 2 3 4
20.	Do my best to follow the instructions for the task	0 1 2 3 4
21.	Decide there is no point in trying to do well	0 1 2 3 4

Attentional Self-Regulation Scale

Instructions: For each of the statements below, please indicate whether or not the statement is characteristic of you or of what you believe. Rate the items using the following scale.

1—not at all true 2 3 4—completely true

1. I can concentrate on one activity for a long time, if necessary.
2. If I am distracted from an activity, I don't have any problem coming back to the topic quickly.
3. If an activity arouses my feelings too much, I can calm myself down so that I can continue with the activity soon.
4. If an activity requires a problem-oriented attitude, I can control my feelings.
5. I can control my thoughts from distracting me from the task at hand.
6. After an interruption, I don't have any problem resuming my concentrated style of working.
7. I stay focused on my goal and don't allow anything to distract me from my plan of action.

Occupational Self-efficacy

Instructions: For each of the statements below, please indicate whether or not the statement is characteristic of you or of what you believe. Rate the items using the following scale.

1 = not at all true 2 3 4 5 6 = completely true

1. I can remain calm when facing difficulties in my job because I can rely on my abilities
2. When confronted with a problem in my job, I can usually find several solutions.
3. Whatever comes my way in my job, I can usually handle it.
4. My past experiences in my job have prepared me well for my occupational future.
5. I meet the goals that I set for myself in my job.
6. I feel prepared for most of the demands in my job.

Mission Orientation: I am committed to the TSA's mission.

Appendix C:
Job Performance Rating Scale (JPRS)

Supervisor Last Name: ___________________________ **Tenure with TSA: _____**

Shift: _______________

Job Performance Rating Scale

Instructions: In this exercise, you will rate your team members' abilities to perform tasks related to behavior detection. Please consider how well each team member has performed the listed behavior detection tasks in the past month. The goal of the scale is to understand how well your team member performs his or her BDA responsibilities. Please rate your team members as accurately as possible.

Your ratings will **NOT** count towards TOPS evaluations or other aspects of the APR; instead the ratings will be used only by researchers to better understand factors that lead to successful behavior detection performance. To ensure your ratings will not be used in any manner other than for research purposes, the names of your team members are not listed on this form. Please refer to the *Employee Name—Participant ID* key to determine who you should rate on each line. No one at TSA or DHS other than you will know how you rate each of your team members. We hope that this assurance will help you give accurate ratings.

On the following pages you will rate performance on these behavior detection dimensions:

- Specific Visual Search Skills
- General Visual Search Skills
- Behavior Detection and Analysis
- General Skills and Abilities

After rating your selected team members on the above behavior detection dimensions, you will rate them on two global performance measures. As with the job dimension ratings, these global ratings will **NOT** count towards TOPS evaluations or other aspects of the APR. These too will be used by researchers only.

Please rate your team member's performance over the past month. For each behavior detection task and global rating, rate every team member on that task or global rating before moving on to the next task or global rating. When rating each team member, consider that person's performance relative to the rating scale, not relative to the performance of the other team members listed. That is, do not compare the performance of the listed team members against each other.

Dimension: Domain Specific Visual Search Skills

Instructions: Please rate your team member's abilities to perform these tasks related to behavior detection and analysis using the following 4-point scale.

Not so well	Somewhat well	Very well	Extremely well
1	2	3	4

Write "NA" if a behavior does not apply

1. Identifies indicators when they are displayed by passengers

	Rating
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________

2. Maintains situational awareness of environment and passengers while conducting BDA

	Rating
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________

3. Maintains focus and attention over the course of BDA operations

	Rating
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________

4. Focuses on relevant passengers' behaviors while ignoring other distracting information or stimuli

	Rating
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________

5. Uses visual observation skills to recognize differences or similarities in passengers and events

	Rating
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________

6. Maintains visual contact with passenger queue

	Rating
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________

Dimension: Domain General Visual Search Skills

Instructions: Please rate your team member's abilities to perform the tasks related to behavior detection and analysis use the following 4-point scale

Not so well	Somewhat well	Very well	Extremely well
1	2	3	4

Write "NA" if a behavior does not apply

1. Performs monotonous tasks without getting distracted

Part ID	Rating
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________

2. Concentrates his/her attention on a task for long periods of time

Part ID	Rating
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________

3. Pays attention to detail

Part ID	Rating
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________

4. Finds things that are hidden

Part ID	Rating
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________

5. Foresees problems before they arise

Part ID	Rating
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________

6. Notices things that seem out of place

Part ID	Rating
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________

Dimension: Domain General Visual Search Skills (Cont.)

Instructions: Please rate your team member's abilities to perform the tasks related to behavior detection and analysis use the following 4-point scale

7. Performs several tasks simultaneously without sacrificing performance	
	Rating
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________

8. Maintains situation awareness of surroundings and environment	
	Rating
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________

Dimension: Behavior Detection and Analysis

Instructions: Please rate your team member's abilities to perform the tasks related to behavior detection and analysis use the following 4-point scale

Not so well	Somewhat well	Very well	Extremely well
1	2	3	4

Write "NA" if a behavior does not apply

1. Establishes an accurate, in-depth environmental baseline by identifying and communicating factors that can affect passenger behavior	Rating
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________

2. Accurately adjusts the environmental baseline to accommodate all changes in typical behaviors and appearances	Rating
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________

3. Accurately and consistently identifies individuals whose behavior or appearance factors deviate from the established environmental baseline	Rating
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________

4. Accurately refers travelers who exceed indicators to additional screening	Rating
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________

5. Engages passengers in varying verbal exchanges while moving through passenger queue in order to observe passenger behavior and reactions	Rating
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________

6. Correctly and continuously positions oneself relative to partner to ensure effective observational coverage while conducting active engagement	Rating
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________
Part ID__________	________________

Dimension: Behavior Detection and Analysis (cont.)

Instructions: Please rate your team member's abilities to perform the tasks related to behavior detection and analysis use the following 4-point scale

Not so well	Somewhat well	Very well	Extremely well
1	2	3	4

Write "NA" if a behavior does not apply

7. Strategically engages individuals exhibiting behavioral indicators.	Rating
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________

8. Utilizes partner to detect behavioral indicators	Rating
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________

Dimension: General Skills and Abilities

Instructions: Please rate your team member's abilities to perform the tasks related to behavior detection and analysis use the following 4-point scale

Not so well	Somewhat well	Very well	Extremely well
1	2	3	4

Write "NA" if a behavior does not apply

1. Acts assertively when necessary

Part ID	Rating
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________

2. Demonstrates sufficient effort on most tasks and assignments

Part ID	Rating
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________

3. Completes written reports in a satisfactory manner

Part ID	Rating
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________

4. Demonstrates confidence, credibility, and professionalism in presence, demeanor, and conduct in performance of duties within the work environment

Part ID	Rating
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________

5. Accomplishes tasks in a thorough and precise manner; double checks the accuracy of information and work products

Part ID	Rating
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________

6. Interacts with passengers in a comfortable manner; approaches others to obtain information with ease

Part ID	Rating
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________
Part ID__________	______________

Global Ratings

Please use the below 1 to 5 rating scale for the next two items

1 Poor	2 Below Average	3 Average	4 Above Average	5 Outstanding

1. Overall, how would you rate each individual's technical performance as a BDO?	
	Rating
Part ID__________	____________
Part ID__________	____________
Part ID__________	____________
Part ID__________	____________

2. Overall, how would you rate each individual's ability to conduct behavior detection?	
	Rating
Part ID__________	__________
Part ID__________	__________
Part ID__________	__________
Part ID__________	__________

Using the below rating scale, please indicate how often in the past month you have had an opportunity to observe the listed BDO. This rating is needed to ensure that you have had ample opportunity to rate this employee's abilities.

1 Almost Never Had the Opportunity	2 Seldom Had the Opportunity	3 Sometimes Had the Opportunity	4 Frequently Had the Opportunity	5 Very Frequently Had the Opportunity

1. Perform behavior detection and analysis activities.	
	Frequency
Part ID__________	____________
Part ID__________	____________
Part ID__________	____________
Part ID__________	____________

2. Demonstrate knowledge of behavioral detection standard operating procedures (SOPs)	
	Frequency
Part ID__________	____________
Part ID__________	____________
Part ID__________	____________
Part ID__________	____________

Appendix D: JPRS Rater Training

Hello, my name is [INSERT NAME} from RTI International, a non-profit research organization. I am part of the project team working with staff at DHS S&T and TSA. We are conducting research to determine a better way of identifying individuals who are more likely to be successful at Behavior Detection. You all are here, as supervisors, to help us with this research. There are 3 points I want to emphasize regarding this work by way of background for what we'll be doing today.

The first point is that our main objective in the project is to evaluate a test, the Behavior Detection Battery, as a predictor of behavior detection performance.

In other words, we want to find out if staff who score highly on the Behavior Detection Battery also tend to perform effectively on behavior detection tasks. Accordingly, to evaluate the battery in this way, we need to have current BD trained TSOs take the battery *and* have each of the same BD trained TSOs assessed on their on-the-job performance of behavior detection. This is where you can help us. We are going to ask you to provide—for research only—ratings of BD trained TSOs who have taken the Behavior Detection Battery. If we determine that there is a relationship between Behavior Detection Battery scores and BD trained TSO performance, the battery may useful in helping select non-BD trained TSOs for BD training.

A second point to emphasize is that the rating scales you will be working with today were developed with considerable help from several TSA staff with knowledge about behavior detection tasks.

Consequently, we believe you will find them quite relevant for evaluating the behavior detection component of TSO performance.

The final point is that our evaluation of the validity of the Behavior Detection Battery can only be accomplished if you provide us with accurate ratings of the BD trained TSOs you will be assessing today. In other words, this is not a popularity contest.

We really need you to be as accurate as possible in your ratings. In this regard, you should know that your ratings will be kept strictly confidential. Our staff will be the only ones who see your ratings, and they will be used only for the research purpose of validating the Behavior Detection Battery.

So, let's get into the rating task. We will now hand out the Job Performance Rating Scale for recording the ratings. When you get the documents, please read the instructions. Take a moment to read the instructions [WAIT].

Let me briefly discuss a couple of the features of the rating scales. Take a look at the first dimension: **Domain Specific Visual Search Skills**.

Immediately below the dimension names are the rating instructions for the dimension.

Immediately below that is the four-point scale you will use to rate that dimension. The scale ranges from 1—'Not so well' on the low end to 4—'Extremely Well' on the high end. If a behavior does not apply to a particular employee, you can write NA. It is important to note that we are only interested in the employee's behavior over the past month.

Below the rating scale are the behavioral statements the BD trained TSOs will be rated on. For each behavioral statement, you rate each of your listed team members on that statement before moving on to the next behavioral statement. The statements flow from left to right as indicated by the arrows.

These ratings forms do not include the names of each employee. For confidentiality reasons, the form only lists a participant's ID. You should each have an *Employee Name—Participant ID* key that shows which of your team members is represented by each participant ID. When completing your ratings, please refer to this key often to make sure you are rating the correct person. Please do not write the name of the officers you are rating on the rating form. Does anyone not have a *Employee Name—Participant ID* key? Does anyone have on their key an employee that is not yours? Does anyone missing an employee on their key?

After rating each team member on the 6 behavioral statements under **Specific Visual Search Skills**, you move on to the next dimension—**General Visual Search Skills**. After rating each team member on the 8 behavioral statements under **General Visual Search Skills,** move to **Behavior Detection and Analysis.**

After going through the four job dimensions, we want you to rate your selected team members on two global ratings. Look at page 8. Notice that the scale at the top of the page is different than the scale used for the job dimensions. The global rating scale is a 5-point scale ranging from 1—'poor' on the low end to 5—'outstanding' on the high end. You'll use this scale for the two questions above the black line.

The final two items ask you how often in the past month you have had the opportunity to observe the listed officers performing behavior detection and analysis, and demonstrating knowledge of behavioral detection standard operation procedures. For these two items, use the scale below the black line. This scale ranges from 1—'almost never had the opportunity' on the low end to 5—'very frequently had the opportunity' on the high end.

For each item, it is important to remember to rate each officer as accurately and as objectively as you can.

Now, before you actually start your ratings there are a couple more things I want to go over with you. When rating the performance of others, *we all* have the tendency to make certain kinds of errors. At this point, I'm going to take a minute to review with you three very common rating errors, so that you will hopefully be able to avoid these when you do your evaluations.

The first error is called HALO ERROR. What this means is that you have a general good or bad impression of the person you're evaluating and this impression tends to influence all of your ratings of him or her. For example, let's say you're rating Officer Smith. You feel this officer is pretty good overall. So you give him fairly high ratings in all of the performance areas. For example, you might give him 4s across all of the statements. Now, it's very

unlikely that any TSO performs all of these job duties extremely well. The reason for this is because each factor is a relatively independent or separate area of the job, and each TSO you are rating is likely to be strong in some areas and weaker in other areas. What we want you to do is tell us about each TSOs *strengths and weaknesses*. In other words, in what areas does the person perform well and in what areas does the person perform less effectively?

The second error that raters sometimes make is to allow things that have nothing to do with job performance influence their ratings. For example, someone's family background or education or previous experience led you to rate the person in certain ways—either high or low. Today, we want you to try to put anything that is not related to *actual* job performance out of your mind and to provide us with the most accurate and *objective* ratings that you can.

The last rating error I want to go over with you is called the SAME-LEVEL-OF-EFFECTIVENESS ERROR. What this means is that raters sometimes tend to give the exact same rating to all of the persons they rate. So, for example, on Behavior Detection and Analysis we might see ratings of "2", "2", "2", "2" across all of the TSOs being evaluated. And on General Visual Search Skills we might see all 4s, and so on. What I'm saying is that we not only want you to tell us about each TSO's strengths and weaknesses, but we also want you to indicate differences *between* TSOs who perform well in each area and those who perform less well in that area.

Now that I've gone through these 3 errors, there's one final point that I want to stress again. Although we don't want you to make rating errors, what's most important is that you rate each of the TSOs accurately. For instance, 3 of the TSOs you're rating may actually perform the same on one of the statements, or you may feel that one of these TSOs actually performs at the "4" level on many of the factors. If this is the case, then by all means, rate the individuals in this way. However, when real differences exist, then your ratings should reflect these differences.

In terms of actually going through the rating process, it will be easiest and most efficient if you rate each of the TSOs on the first statement in the first dimension, and then rate each TSO on the second statement in the first dimension. Follow this procedure for each of the remaining statements and dimensions, evaluating each TSO on a statement before turning to the next statement under that dimension.

All right, you should now be well-prepared to make your ratings. Thank you for your help with this important research. If you haven't done so already, please write your tenure and shift on the first page. Please begin making your ratings starting on page 2.

Appendix E:
Behavior Detection Battery Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Attentional Self-Regulation	199	21	7	28	24.74	3.593	12.909	-1.232	0.172	2.028	0.343
Boredom Proneness	199	18	0	18	7.0603	3.45628	11.946	0.606	0.172	0.014	0.343
CITS—Avoidance-Focused	150	14	7	21	10.94	3.60319	12.983	0.739	0.198	-0.176	0.394
CITS—Emotion-Focused	150	17	7	24	11.0333	4.30766	18.556	1.032	0.198	0.056	0.394
CITS—Task-Focused	150	27	8	35	26.76	5.1614	26.64	-0.853	0.198	0.895	0.394
Commitment to TSA	199	5	1	6	5.71	0.831	0.69	-3.904	0.172	17.031	0.343
Conscientiousness	198	78	22	100	86.1263	11.32399	128.233	-2.013	0.173	6.758	0.344
Job Boredom	196	94	17	111	52.4337	17.40471	302.924	0.696	0.174	0.68	0.346
Luggage Image Search Task *Accuracy*	148	54	43	97	75.4189	12.18397	148.449	-0.346	0.199	-0.71	0.396
Luggage Image Search Task *Detection Sensitivity*	148	3.83	-0.14	3.68	1.656	0.80428	0.647	0.116	0.199	-0.363	0.396
Luggage Image Search Task *Response Bias*	148	2.02	-0.98	1.04	-0.0594	0.40445	0.164	0.232	0.199	0.126	0.396
Need for Cognition	200	50	38	88	66.3	10.68272	114.121	-0.155	0.172	-0.306	0.342
Occupational Self-Efficacy	199	35	7	42	38.73	4.347	18.896	-2.685	0.172	13.693	0.343
O-SPAN—Working Memory	157	73	2	75	48.58	15.756	248.258	-0.652	0.194	-0.198	0.385
Simulation-Based Behavior Detection Task	155	72.34	18.75	91.09	59.0305	16.56554	274.417	-0.442	0.195	-0.308	0.387
SSSQ—Distress	151	22	8	30	11.4901	4.26125	18.158	2.078	0.197	4.979	0.392
SSSQ—Engagement	151	23	17	40	32.2715	5.04768	25.479	-0.514	0.197	-0.201	0.392
SSSQ—Worry	151	26	8	34	16.404	6.11793	37.429	0.795	0.197	0.262	0.392

(continued)

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
SVT—Simultaneous Condition *Decrement*	72	3.64	-1.84	1.8	0.2945	0.78006	0.608	-0.027	0.283	-0.222	0.559
SVT—Successive Condition *Accuracy*	64	0.249	0.75	0.998	0.95236	0.063408	0.004	-2.006	0.299	3.196	0.59
SVT—Successive Condition *Detection Sensitivity*	68	3.56	0.81	4.37	3.4867	0.88453	0.782	-1.286	0.291	1.041	0.574
SVT—Simultaneous Condition *Accuracy*	72	0.734	0.263	0.997	0.90481	0.170738	0.029	-2.937	0.283	7.609	0.559
SVT—Simultaneous Condition *Detection Sensitivity*	74	4.55	-0.24	4.31	3.1424	1.12948	1.276	-1.678	0.279	2.336	0.552
SVT—Simultaneous Condition *Response Bias*	74	2.18	-1.33	0.86	0.2142	0.37536	0.141	-2.33	0.279	7.04	0.552
SVT—Successive Condition *Decrement*	63	3.32	-1.52	1.8	0.3136	0.63764	0.407	-0.123	0.302	0.973	0.595
SVT—Successive Condition *Response Bias*	68	1.64	-0.26	1.38	0.2924	0.3032	0.092	1.393	0.291	2.221	0.574
Video-Based Passenger Observation Task	190	90.91	9.09	100	59.1977	17.74164	314.766	-0.387	0.176	-0.129	0.351

Appendix F: Power Analysis

We conducted a power analysis to determine the minimum number of BDOs to include in the sample to a find meaningful effect between performance on the predictor battery and criterion measures. However, rather than calculating the minimum sample size for a given experimental or quasi-experimental design, we set the desired sample size to a specific number and then examined how clustering effects may impact the minimum detectable effect sizes (MDES).

The MDES is the smallest meaningful effect (in this case, the smallest association) one can expect to detect for a specific sample size. We chose this method over the more traditional approach of calculating the minimum because there is less variation than would be expected from a simple random sample.

To guide our analysis, we assumed a sample size of 300 BDOs (~20 from each airport), an observed power level of .80, and a type 1 error rate (alpha) of .05. Table F1 shows the impact of clustering on the effective sample size (ESS) and MDES. Clustering is measured using the intraclass correlation (ICC) coefficient. Higher coefficients represent high clustering. The ESS indicates how many effective participants are left in the sample as a result of clustering.

Table F1. Proposed Sample Size

		Minimum Detectable Effect Size (MDES)		
			Odds Ratio (OR)	
ICC	Effective Sample Size (ESS)	Correlation (*r*)	Outcome Proportion = .2	Outcome Proportion = .5
0.00	300	*r* = 0.16	OR = 1.5	OR = 1.4
0.05	147	*r* = 0.23	OR = 1.8	OR = 1.6
0.10	98	*r* = 0.29	OR = 2.0	OR = 1.8
0.15	73	*r* = 0.32	OR = 2.3	OR = 1.9

As shown, with an ICC of 0.00 and an effective sample size (ESS) of 300 BDOs, the minimum detectable relationship between two variables (Rxy) is r = .16. That is, a correlation value of .16 between a predictor and criterion would be considered a meaningful relationship. As the ICC increases, the effective sample size decreases and a stronger relationship between two variables is needed to detect a significant effect.

In addition to examining the MDES for correlations we also calculated the sample size for the TSO population to facilitate analysis of group differences on the selection measures. Results of our analysis showed that to achieve statistical power of .80, a medium effect size, and a type 1 error rate of .05 (alpha), we would need to include an additional 76 TSOs in our sample. This would allow us to test for meaningful differences between BDOs and TSOs on the selection battery (Keppel & Wickens, 2004).

Appendix G: Correlations between Individual Characteristics and Behavior Detection Visual Search Performance for TSOs and BDOs

Measures of Comparison	TSOs	BDOs
O-SPAN—Working Memory compared to		
Video-Based Passenger Observation Task	r(76) = .34, p = .002	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	ns	r(64) = .24, p = .048
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	r(28) = .42, p = .022
Shortened Vigilance Task—Successive Condition Accuracy	r(34) = .61, p < .001	r(26) = .48, p = .011
Shortened Vigilance Task—Successive Condition Detection Sensitivity	r(35) = .63, p < .001	r(29) = .36, p = .046
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
Conscientiousness compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	ns	ns
Luggage Image Search Task Detection Sensitivity	r(78) = -.24, p = .036	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns

(continued)

Measures of Comparison	TSOs	BDOs
Need for Cognition compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	r(64) = .32, p = .010
Luggage Image Search Task Accuracy	ns	r(64) = .25, p = .040
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	r(26) = .47, p = .013
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
Boredom Proneness compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	r(64) = -.37, p = .003
Luggage Image Search Task Accuracy	ns	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
Job Boredom compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	r(77) = .22, p = .050	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns

(continued)

Measures of Comparison	TSOs	BDOs
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	r(39) = .33, p = .036	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
Attentional Self-Regulation compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	ns	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	r(26) = .42, p = .027
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
Occupational Self-Efficacy compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	ns	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns

(continued)

Measures of Comparison	TSOs	BDOs
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
Commitment to TSA Mission compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	ns	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	$r(34) = .39, p = .018$	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
SSSQ—Distress compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	ns	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
SSSQ—Engagement compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	ns	ns
Luggage Image Search Task Detection Sensitivity	ns	ns

(continued)

Measures of Comparison	TSOs	BDOs
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
SSSQ—Worry compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	ns	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
CITS—Task-Focused compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	$r(62) = .32, p = .011$
Luggage Image Search Task Accuracy	$r(75) = .31, p = .007$	ns
Luggage Image Search Task Detection Sensitivity	$r(75) = .35, p = .002$	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	$r(39) = .36, p = .022$	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	$r(29) = .44, p = .014$
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns

(continued)

Measures of Comparison	TSOs	BDOs
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
CITS—Emotion-Focused compared to		
Video-Based Passenger Observation Task	r(76) = -.30, p = .009	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	ns	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	r(75) = .24, p = .037	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	r(33) = -.37, p = .030	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
CITS—Avoidance-Focused compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	ns	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	r(39) = -.33, p = .038	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	r(39) = -.32, p = .040	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	r(33) = .36, p = .033	ns

Appendix H:
Correlations between Demographics and Behavior Detection Visual Search Performance for TSOs and BDOs

Measures of Comparison	TSOs	BDOs
TSA Tenure compared to		
Video-Based Passenger Observation Task	r(106) = -.22, p = .021	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	r(79) = -.27, p = .013	r(64) = -.38, p = .002
Luggage Image Search Task Detection Sensitivity	r(79) = -.25, p =.022	r(64) = -.40, p = .001
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	r(25) = -.40, p = .038
Shortened Vigilance Task—Successive Condition Detection Sensitivity	r(35) = -.36, p = .028	r(28) = -.42, p = .021
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
Pay Grade compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	ns	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns

(continued)

Measures of Comparison	TSOs	BDOs
Position Tenure compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	r(107) = -.24, p = .012	ns
Luggage Image Search Task Accuracy	ns	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
Age compared to		
Video-Based Passenger Observation Task	r(106) = -.27, p = .005	r(79) = -.23, p = .039
Simulation-Based Behavior Detection Task	r(85) = -.32, p = .003	r(64) = -.35, p = .004
Luggage Image Search Task Accuracy	r(78) = -.38, p = .001	r(64) = -.37, p = .002
Luggage Image Search Task Detection Sensitivity	r(78) = -.40, p < .001	r(64) = -.43, p < .001
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	r(39) = -.37, p = .018	r(30) = -.38, p = .031
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns

(continued)

Measures of Comparison	TSOs	BDOs
Education compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	ns	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
Prior military experience compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	$r(79) = -.27, p = .017$	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	$r(39) = -.35, p = .027$	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
Prior military visual search experience compared to		
Video-Based Passenger Observation Task	ns	$r(21) = -.50, p = .015$
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	$r(15) = -.54, p = .027$	ns

(continued)

Measures of Comparison	TSOs	BDOs
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
Prior law enforcement experience compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	ns	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	$r(30) = -.35, p = .048$
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
Prior law enforcement visual search experience compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	ns	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns

(continued)

Measures of Comparison	TSOs	BDOs
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
Frequency of examining X-ray images compared to		
Video-Based Passenger Observation Task	$r(107) = .20, p = .038$	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	$r(79) = .28, p = .012$	ns
Luggage Image Search Task Detection Sensitivity	$r(79) = .29, p = .009$	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	$r(39) = .32, p = .040$	ns
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	$r(25) = -.40, p = .037$
Frequency of examining AIT images compared to		
Video-Based Passenger Observation Task	$r(107) = .22, p = .022$	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	$r(79) = .33, p = .002$	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	$r(29) = .42, p = .020$

(continued)

Measures of Comparison	TSOs	BDOs
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns
Frequency of examining passengers as part of BD compared to		
Video-Based Passenger Observation Task	ns	ns
Simulation-Based Behavior Detection Task	ns	ns
Luggage Image Search Task Accuracy	ns	ns
Luggage Image Search Task Detection Sensitivity	ns	ns
Luggage Image Search Task Response Bias	ns	ns
Shortened Vigilance Task—Simultaneous Condition Accuracy	ns	ns
Shortened Vigilance Task—Simultaneous Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Simultaneous Condition Response Bias	$r(40) = -.32$, $p = .038$	ns
Shortened Vigilance Task—Simultaneous Condition Decrement	ns	$r(29) = -.49$, $p = .016$
Shortened Vigilance Task—Successive Condition Accuracy	ns	ns
Shortened Vigilance Task—Successive Condition Detection Sensitivity	ns	ns
Shortened Vigilance Task—Successive Condition Response Bias	ns	ns
Shortened Vigilance Task—Successive Condition Decrement	ns	ns