



The Impact of Information Presentation on Visual Inspection Performance in the International Nuclear Safeguards Domain

Laura E. Matzen^(✉), Mallory C. Stites, Heidi A. Smartt,
and Zoe N. Gastelum

Sandia National Laboratories, Albuquerque, NM 87123, USA
{lematze, mcstite, hasmart, zgastel}@sandia.gov

Abstract. International nuclear safeguards inspectors are tasked with verifying that nuclear materials in facilities around the world are not misused or diverted from peaceful purposes. They must conduct detailed inspections in complex, information-rich environments, but there has been relatively little research into the cognitive aspects of their jobs. We posit that the speed and accuracy of the inspectors can be supported and improved by designing the materials they take into the field such that the information is optimized to meet their cognitive needs. Many in-field inspection activities involve comparing inventory or shipping records to other records or to physical items inside of a nuclear facility. The organization and presentation of the records that the inspectors bring into the field with them could have a substantial impact on the ease or difficulty of these comparison tasks. In this paper, we present a series of mock inspection activities in which we manipulated the formatting of the inspectors' records. We used behavioral and eye tracking metrics to assess the impact of the different types of formatting on the participants' performance on the inspection tasks. The results of these experiments show that matching the presentation of the records to the cognitive demands of the task led to substantially faster task completion.

Keywords: Visual inspection · Situational awareness ·
Information visualization

1 Introduction

International nuclear safeguards are measures that provide assurance to the global community that nations are using nuclear technologies for peaceful purposes. The International Atomic Energy Agency (IAEA), which operates under the auspices of the United Nations, is the agency tasked with verifying States' safeguards agreements. A State declares nuclear materials and facilities and the IAEA periodically verifies those declarations to ensure that nuclear materials are not being diverted from known (safeguarded) facilities and that safeguarded facilities are not being misused for undeclared nuclear purposes. They also attempt to detect any undeclared nuclear activities within a State.

The basic verification method used by the IAEA is nuclear material accountancy (NMA), which is achieved through nuclear materials measurements and examination of records and reports. The IAEA also inspects nuclear facilities to determine operational status, design, and production capacity. Containment and surveillance technologies, such as seals and cameras, are applied to maintain continuity of knowledge for nuclear materials between inspection intervals.

During a facility inspection, IAEA inspectors complete tasks such as verifying that seals have not been tampered with, verifying the inventory of nuclear material by checking seal numbers against the IAEA and facility records, comparing the facility's records to their declarations to the IAEA, taking material measurements, and looking for any anomalies in a facility that may indicate misuse. Safeguards inspections are physically and cognitively demanding [1]. The inspectors are working under time pressure in an industrial environment that may be loud, hot, and/or cramped, in addition to containing radiological hazards. For many aspects of the inspections they must wear protective gear which makes it more difficult to manipulate the tools they need to take samples and record observations. They may not share a common language with the facility operators and are often dealing with jet lag on top of the demands of the working environment. While working in this challenging environment, the inspectors must take care to record accurate data and notes, while also being on the alert for any subtle discrepancies or indications of unusual activity in the facility.

Despite the importance of an IAEA inspector's role and the cognitive demands of the job, there has been very little application of cognitive science research to this domain [2]. To address this gap, our research team conducted an evaluation of key safeguards inspection tasks to identify where cognitive science research methods could be applied to support the inspectors' cognitive processing in the field [3]. One of the areas that we identified was visual inspection. IAEA inspectors must complete several types of visual inspection tasks, including paper-based tasks, such as comparing the facility's inventory and shipment records to State declarations and records from prior inspections, and object-based tasks, such as finding seal numbers on containers of nuclear materials and checking them against IAEA records.

There has been a great deal of research on visual search in general [4], as well as visual search of lists [5, 6] and visual inspection in industrial contexts [7–12]. These studies have shown that factors such as the work environment [9, 12], task structure [9, 11], feedback [7], list formatting [5] and other types of job aids [8] impact visual search and inspection performance. Although safeguards inspectors do not have control over many of these factors, it may be feasible to format their own records and inspection-related materials in ways that make their visual inspection tasks faster and easier. We conducted a series of experiments to test the impact of changing the formatting of the inspector's materials on the speed and accuracy of their inspection performance.

In our first study in this area [13], we developed a computer-based mock inspection task. Participants saw two lists of seal and container numbers displayed side-by-side on the screen. One of these lists was designated the "inspector's list" and one was the "facility's list." Participants were tasked with checking all the items on the inspector's list against the facility's list and marking which seals were present, which were missing, and which (if any) were anomalous in other ways. While the facility's list was always presented in a random order, we altered the presentation of the inspector's list

by changing its order and color coding, two factors long known to impact visual search performance [5, 14–17]. The experiment found that participants had equally high accuracy across all of the list presentation conditions, but very different response times. Participants were fastest when the order of the seals on the inspector’s list matched the order of the seals on the facility’s list. When the order did not match, participants benefited from having color coding to narrow their search of the facility’s list, which significantly improved their response times.

Given the results of this study, we expanded this line of research to other types of inspection tasks. In [13], participants were tasked with verifying all of the seals on the facility’s list. However, in real-world IAEA inspections, it is common for inspectors to check a randomly determined, statistically representative subset of the seals in a facility during each inspection. Checking a subset of the full list may change how the inspectors use the lists, which may in turn impact which list presentation conditions lead to the biggest benefits to the inspectors’ performance. The work presented here addresses this scenario.

Experiment 1, like our prior study, involved a list-to-list comparison activity, except that participants were only looking for a subset of the seals in the facility’s list. Experiments 2 and 3 represented inspection activities in which the inspectors must walk through a facility to check a list against physical items, such as sealed containers. Although these were computer-based tasks, they were designed to mimic list-to-item inspection activities. In Experiment 2, participants had a list of seals to verify but could only view one sealed container at a time, mirroring the process of checking a subset of the sealed containers in one room of a facility. In Experiment 3, participants used a map to navigate between different “rooms” in a facility. Across all three experiments, the information provided to the participants was manipulated to determine the impact of list order and different types of information about a seal’s likely location on the speed and accuracy of the inspection.

2 Experiment 1

In Experiment 1, participants were given a mock safeguards inspection task in which they were asked to compare two lists to ensure that the information matched. The “inspector’s list” contained a subset of half of the items on the “facility’s list,” mimicking a paper-based inspection task in which inspectors verify a representative subset of a facility’s records. The order and color coding of the inspector’s list was manipulated across six conditions and participants were assessed in terms of their accuracy and response times for each condition. Eye tracking data were collected to identify any differences in inspection strategy across the list presentation conditions.

2.1 Method

Participants. Nineteen participants were recruited from the employee population of Sandia National Laboratories and were compensated for their time. Four participants were later excluded from the analysis due to dropped or noisy eye tracking data. The

remaining 15 participants (10 female) had an average age of 32 years. Four of the participants held a high school degree, three held a bachelor's degree, four held a master's degree, and three held a PhD.

Materials. As in our prior study [13], the experimental materials consisted of six sets of lists containing seal numbers and container numbers. The inspector's list, presented on the left side of the computer screen, contained 18 pairs of seal and container numbers arranged in two columns. The facility's list, presented on the right side of the screen, contained 36 pairs of seal and container numbers arranged in four columns. See Fig. 1 for an example.



Fig. 1. Example of the screen layout used in Experiment 1. This example shows the color-coded facility order condition. (Color figure online)

The seal numbers were six-digit numerical strings. Within each condition, the first digit of the seal number was always the same and the five final digits were pseudorandomly generated such that every digit (0–9) appeared approximately the same number of times in each position. This was done to avoid any patterns within the seal numbers that could have made some numbers more memorable than others. The container numbers consisted of two letters and two numbers, separated by a hyphen, such as “AB-37.” Each container number was unique, although the same letter pairs appeared in multiple container numbers.

On the facility's list, there were 20 filler items that did not appear on the inspector's list. The remaining items corresponded to the conditions outlined in Table 1. Some of the conditions contained transposed digits, which were intended to make the inspection more difficult. If participants did not pay close attention, they could mismatch or

mis-categorize these items, which could lead to confusion later in the inspection process. The types of transpositions were the same as in [13].

Table 1. Conditions and example items for Experiment 1.

Condition	Number of items	Examples	
		Inspector's list version	Facility list version
Match	8	845432 EF-36	845432 EF-36
Wrong container	2	864413 VZ-97	864413 NP-19
Missing	2	894320 AB-41	N/A
Transpose	2	847186 GI-82	847168 GI-82
Transpose match	4	835983 EF-34	835983 EF-34
		835893 CD-57	835893 CD-57

The experiment consisted of six inspection tasks. The seal-container pairs in the facility's list were always presented in a random order, but the presentation of the information on the inspector's list was manipulated across the six blocks. The items on the inspector's list appeared in one of three orders: random order (fixed so that it was the same order for all participants), numerical order, or facility order (in which the seals were presented in the same order as those in the facility's list). There were also two color-coding conditions. In half of the blocks, all the list items were presented in black font. In the other half, each column of the facility's list was assigned a color and the items on the inspector's list were color-coded according based on which column on the facility's list contained the corresponding seal-container pair. The ordering and color-coding conditions were fully crossed, creating a 3×2 within-subjects design.

During each block, the background color of the screen changed two to three times. The possible colors were purple, blue, and teal (examples of the colors are shown in Figs. 1, 4 and 6). The changes were linked to specific seals, such that after a participant clicked on that seal, the background color would change on the next trial. The seals that triggered the color changes were different for each block. The color change detection task was included to encourage participants to maintain their situational awareness by attending to a secondary task while completing their primary inspection task. This mimics the work of safeguards inspectors, who must maintain their overall situational awareness in addition to completing their inspection tasks. There were relatively few color changes per block, so the participants' accuracy on the color change detection task was not analyzed separately for the different inspection conditions.

Procedure. After giving their informed consent, participants were seated in a dimly lit, sound attenuating booth so that their eyes were 80 cm from the computer monitor. Participants completed a practice session that explained the task and allowed them to complete a shortened version of the inspection. After the practice block, the eye tracker was calibrated. Eye tracking data was collected with a Fovio eye tracker and recorded and analyzed with EyeWorks software. The participants completed a five-point calibration sequence, and then the accuracy of the calibration was assessed by the

experimenter and repeated if necessary. The calibration process was repeated prior to each block.

The participants completed the six blocks of the experiment in a counterbalanced, pseudorandom order. Each block began with a description of how the inspector's list would be organized. The participants were instructed to check off each item on the inspector's list. When they clicked on an item in the inspector's list, four response choices appeared in the center of the screen. The choices were "Seal present, correct container," "Seal present, incorrect container," "Seal missing" and "Other issue." Participants clicked on one of the four choices to indicate their response for that seal. After a response was recorded for a seal, that seal was grayed out on the inspector's list to indicate that it had been checked off. Following each response, a fixation cross was presented in the center of the screen for 1.5 s, initiating the next trial.

Participants were instructed to click on a button labeled "Color Change" as soon as they noticed a change in the background color. Clicking on the "Color Change" button also initiated a new trial. Once participants had checked off all the seals on the inspector's list, they clicked the "Inspection Complete" button. Then they were asked how many times the background color changed during the inspection task. Their choices ranged from zero to four.

Finally, participants were asked to describe their search strategy. The instructions stated: "Please give a brief description (2–3 sentences) of the strategy that you used for this inspection task. For example, which list did you start from? What visual cues were you looking for? Did you ever switch to a different strategy during the task, and if so, what was it?" The participants typed their answer in a text box on the screen. Upon finishing each inspection task, participants were given a short break. In total, the experiment lasted 1–1.5 h, depending on how fast the participants completed each task.

2.2 Results

Accuracy. Across all blocks, the participants detected an average of 61% ($SD = 41\%$) of the background color changes in real time and 79% ($SD = 21\%$) of the color changes when asked at the end of the inspection to report the total number of changes that had occurred. The participants were not penalized for overestimating the number of color changes. These results indicate that the participants generally maintained their awareness of the secondary task.

For the primary task, the seal checking task, participants performed near ceiling on all inspection conditions, correctly identifying which seals were present, missing, or paired with the wrong container number. For items in the Transpose condition, the responses "Seal Missing" and "Other Issue" were both counted as correct. The average percentage of seals categorized correctly ranged from 95% to 97%. A 3×2 repeated measures ANOVA showed that there were no significant main effects or interactions for the different list presentation conditions (all $F_s < 1$).

Response Times. In contrast with the accuracy results, the participants had very different response times across the six inspection conditions. For each trial, the participants' response time was calculated as the time from trial onset to the time the participant clicked

on one of the seals on the inspector's list. The average response times across all trials for each condition are shown in Fig. 2. A 3×2 repeated measures ANOVA showed that there was a significant main effect of list order ($F(2,70) = 25.42, p < 0.001$), a significant main effect of color coding ($F(1,70) = 41.25, p < 0.001$), and a significant interaction between the two ($F(2,70) = 11.67, p < 0.001$).

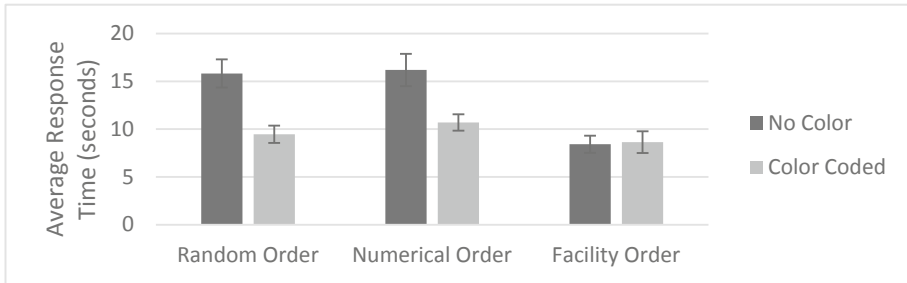


Fig. 2. Average response time for each trial for each of the six conditions in Experiment 1. Error bars represent the standard error of the mean in all figures.

Post-hoc paired t-tests showed that participants responded significantly faster when there was color coding in the random ($t(14) = 7.24, p < 0.001$) and numerical order conditions ($t(14) = 3.55, p < 0.01$). Color coding did not have a significant effect in the facility order condition ($t(14) = 0.21$). Paired t-tests were also used to compare across the list order conditions. When the lists did *not* have color coding, participants were significantly faster for the facility order condition than for the numerical order ($t(14) = 6.58, p < 0.001$) and random order conditions ($t(14) = 7.48, p < 0.001$). The numerical and random order conditions did not differ significantly from one another ($t(14) = 0.53$). For the conditions with color coding, there were no significant differences in response times across the three list order conditions (all t s < 1.69 , all p s > 0.11).

Eye Tracking Data. The eye tracking data showed that the length of the participants' visual search process was the driving factor behind the differences in response times across the six inspection conditions. The average number of fixations per trial is shown in Fig. 3. A 3×2 repeated measures ANOVA showed that there was a significant main effect of list order ($F(2,70) = 11.09, p < 0.001$), a significant main effect of color coding ($F(1,70) = 23.00, p < 0.001$), and a significant interaction between the two ($F(2,70) = 3.23, p < 0.05$).

The gaze data were used to determine how many items the participants scanned as they were searching for each seal number. The seal-container pairs on both lists were labeled as regions of interest (ROIs) and we calculated the average number of ROIs containing gaze data points on each trial. These data are also shown in Fig. 3. Once again, there was a significant main effect of list order ($F(2,70) = 14.25, p < 0.001$), a significant main effect of color coding ($F(1,70) = 30.98, p < 0.001$), and a significant interaction ($F(2,70) = 8.41, p < 0.01$). The patterns seen in both eye tracking analyses mirrored the pattern observed in the response time data.

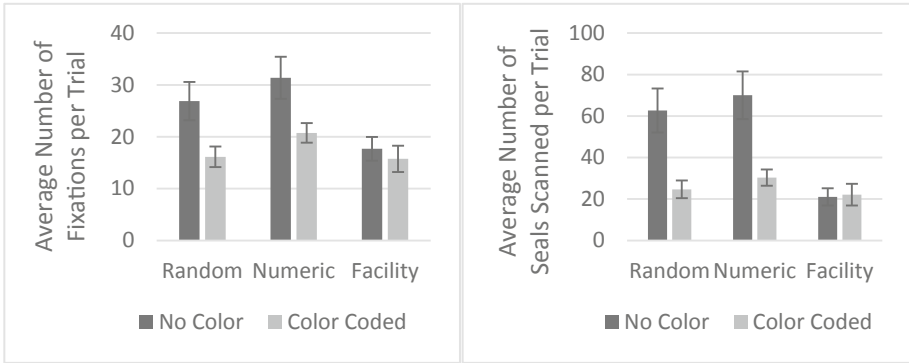


Fig. 3. Average number of fixations (left) and average number of ROIs containing gaze data (right) per trial for each of the six conditions in Experiment 1.

Search Strategy. After completing each inspection task, participants were asked to describe the search strategy that they had used during the task. Across all inspection conditions, all of the participants reported that they started from the inspector’s list and searched for the items in the facility’s list. When there was color coding available, all 15 participants reported using the colors to constrain their search to the appropriate column on the facility’s list. For example, one participant wrote “I looked at the color on my list and then found the right column, then scanned for the first two numbers and then if I found the first two numbers matching I looked to see if the rest matched.”

When the inspector’s list was in the same order as the facility’s list, but with no color coding, 13 of the 15 participants reported that they used the matching order to constrain their search. For example, one participant reported “Since the lists were in the same order, I simply looked at the number on the left and looked to see if there was a match on the right side, below the previous match.” The other two participants did not specify whether they used this information.

When the inspector’s list was in numerical order, participants could have used the order information to constrain their search. If they started from the facility’s list, they could have used the numerical ordering of the inspector’s list to quickly match or eliminate seals. However, none of the participants reported using this strategy, nor did the behavioral or eye tracking results indicate that any participants used this strategy. Only one participant mentioned the numerical ordering, saying “It was really not helpful for me to have my list in number order, having the facility list in number order would have been better.”

2.3 Discussion

The results of the partial-to-full list comparison in Experiment 1 were generally consistent with those for the complete list-to-list comparison used in our prior work [13]. Participants were equally accurate across all of the list presentation conditions, but they had faster response times when their list was ordered to match the order of the items on the facility’s list. The use of color coding to narrow the participants’ search space to

specific columns in the facility's list also had a significant impact on response times. All the participants took advantage of this cue, which allowed them to search more efficiently. This was reflected in faster response times, fewer fixations per trial, and fewer items scanned per trial for the conditions that used color coding.

Interestingly, none of the participants used the numerical ordering to narrow their search space. In our prior work, we found that few participants took advantage of the numerical ordering, but those that used it were able to complete the inspections faster [13]. In the present study, all of the participants searched by choosing an item on the inspector's list and searching for it in the facility's list, regardless of list presentation condition. This is most likely due to the imbalance in the length of the lists in the present study. If participants started from the facility's list, they would need to spend time eliminating the seals that were not on their checklist. Taking advantage of the numerical ordering would have made the search process more efficient but would have increased the number of items that participants needed to search for.

3 Experiment 2

In Experiment 2, we built on the findings of Experiment 1 by extending the experimental paradigm to another aspect of safeguards inspections. In addition to comparing inventory lists to one another, safeguards inspectors must also physically check the seals on containers to verify their presence within a facility and to ensure that the seals have not been tampered with. In this case, inspectors must navigate through a facility to find and check the seals on their list. Unlike the list-to-list comparisons, where all of the information can be placed side-by-side, in this list-to-seal comparison, inspectors are only able to look at one seal at a time.

This scenario changes the dynamics of search process, which may also change which types of list presentation conditions are most helpful. Thus, in Experiment 2 the inspector's lists were the same as in Experiment 1, but instead of checking the inspector's list against a second list, the participants checked the list against images of sealed containers which could only be viewed one at a time. In this scenario, we hypothesized that participants would start by looking at the image of the sealed container, then search for the seal or container number in their list. Searching in this manner should lead more participants to take advantage of non-color cues on the inspector's list, such as numerical ordering.

3.1 Method

Participants. Twelve participants were recruited from the employee population of Sandia National Laboratories and were compensated for their time. The participants (4 female) had an average age of 41 years. One of the participants held a high school degree, one held an associate degree, five held bachelor's degrees, and five held master's degrees. One participant reported a diagnosis of colorblindness.

Materials and Procedure. The seal-container pairs were the same as those used in Experiment 1 and included all of the same seal conditions (Match, Wrong Container,

Missing, Transpose, and Transpose Match). As before, the inspector's list, containing 18 items, was presented in two columns on the left side of the computer screen. Participants checked the list against 36 sealed containers in the "facility." The right side of the screen showed an olive green square representing a container and a yellow circle representing a seal. Buttons labeled "Previous" and "Next" allowed participants to reveal the sealed containers in sequence. Below the container, a text box tracked which seal-container pair was being shown (i.e., "Seal 9/36"). When participants clicked on the buttons to move to another container, there was a short delay (500 ms) before the new container was displayed. This delay simulated physically moving to see the next seal. An example of the screen layout is shown in Fig. 4.

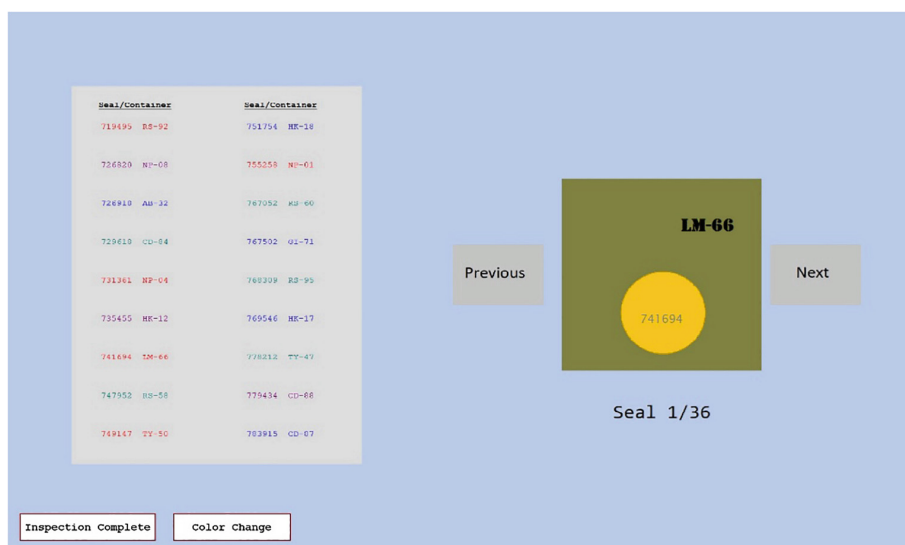


Fig. 4. Example of the screen layout used in Experiment 2. This example shows the color-coded numerical order condition. The box and circle on the right side of the screen represent a container and seal. (Color figure online)

The experiment consisted of six blocks and used the same six variants of the inspector's list that were used in Experiment 1. In this case, the color coding corresponded to the order in which the seals appeared in the facility. Participants were instructed that seals 1–9 in the facility would be presented in red text on the inspector's list, seals 10–18 in teal, seals 19–27 in blue, and seals 28–36 in purple.

The procedure was the same as in Experiment 1, except that eye tracking was not used for this experiment.

3.2 Results

Accuracy. Across all blocks, the participants detected an average of 73% ($SD = 33\%$) of the background color changes in real time and reported an average of 75% ($SD = 33\%$) of the changes when asked for the total number that had occurred during each inspection. There was one participant who did not report seeing any color changes. For the primary task, the participants performed near ceiling on all inspection conditions. The average percentage of seals categorized correctly ranged from 93% to 97%. A 3×2 within-subjects ANOVA showed that there were no significant main effects (all $F_s < 1$), nor was there a significant interaction ($F(2,55) = 2.94, p = 0.06$).

Response Times. The participants' response times were calculated as the time from trial onset to the time the participant clicked on one of the seals on the inspector's list. The average response times across all trials for each list presentation condition are shown in Fig. 5. A 3×2 repeated measures ANOVA showed that there was a significant main effect of list order ($F(2,55) = 7.63, p < 0.01$), but there was not a significant main effect of color coding ($F(1,75) = 0.52$), or a significant interaction ($F(2,55) = 2.33, p = 0.11$).

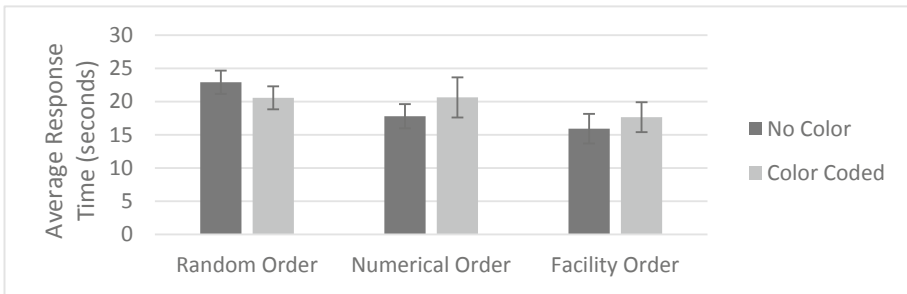


Fig. 5. Average response time for each trial for each of the six conditions in Experiment 2.

Post-hoc paired t-tests were used to compare the list order conditions, collapsed across color coding conditions. The average response time per trial of 21.9 s ($SD = 5.7$) for the random order conditions was significantly longer than the average response time of 19.1 s ($SD = 7.4$) for the numerical order conditions ($t(11) = 2.20, p < 0.05$) and 16.7 s ($SD = 6.6$) for the facility order conditions ($t(11) = 4.10, p < 0.01$). The difference between the numerical order and facility order conditions was not significant ($t(11) = 1.94, p = 0.08$).

Search Strategy. Eleven of the twelve participants stated that they searched by looking at the seal or container number on the right side of the screen, then scanning the inspector's list for that item. The twelfth participant did not specify how s/he searched.

In contrast to Experiment 1, nine participants reported that they used the numerical ordering to constrain their search in one or both numerical order conditions. Five of the

participants reported using the color coding to constrain their search, but two other participants complained that it was distracting and difficult to remember. Six participants reported that they used the ordering of the inspector's list to constrain their search in the facility order condition. Interestingly, in the condition with numerical ordering and color coding, five participants mentioned using the order to help them, while only one participant mentioned the color coding. The rest of the participants did not specify which constraint they used in that condition, if any.

3.3 Discussion

In Experiment 2, we confirmed our hypothesis that a list-to-item comparison would lead to different search strategies than a list-to-list comparison. When participants could only view one seal at a time, they started with the seal and compared it to their list rather than starting from their list as they did in Experiment 1. Both experiments had twice as many items in the "facility" as on the inspector's list. In Experiment 1, this prevented the participants from starting from the facility's list, even when doing so would have been more efficient, as in the numerical order conditions. In Experiment 2, the extra effort required to move between seals flipped the direction of the comparison.

The change in the direction of the participants' search led them to use the cues provided by the inspector's list in different ways. In contrast to Experiment 1, the participants used the numerical ordering to their advantage, leading to performance that was quite similar across the numerical order and facility order conditions. On the other hand, the color coding was not as helpful for this type of search. The analysis showed no main effect of color coding, and on average the participants were numerically slower in the numerical order and facility order conditions when there was color coding. This indicates that the color coding was neutral at best and distracting at worst.

4 Experiment 3

Experiments 1 and 2 indicated that participants benefit from list organization conditions that support their visual search process. In Experiment 3, we tested two additional types of information that could support the search process. First, we replaced the color coding with room numbers. In both of the prior experiments, the color coding provided information about the expected location of each seal. In Experiment 1, the color coding told participants which column of a list would contain the seal (if present), and participants successfully used that information to constrain their searches. In Experiment 2, the color coding told participants which group of seals (1–9, 10–18, etc.) would contain the items from their list. In this case, the participants were beginning their search from the seals rather than from their list and did not derive much benefit from the color coding. They may also have found the color coding more difficult to interpret or to remember. In Experiment 3, we used room numbers rather than color coding to provide location information. The 36 seal-container pairs in the "facility" were divided into four different "rooms." In some of the inspection conditions, the inspector's list included room numbers to indicate which room should contain each seal-container pair. Unlike the color coding used in the other experiments, which was always accurate for

seals that were present in the facility, the room number list was sometimes incorrect. Our goal was to assess the impact of a different method for providing location information as well as the impact of occasional inaccuracies in that information.

In addition to the presence or absence of room numbers, we also manipulated the way in which the participants moved from one container to another. To move between the four “rooms” in the facility, the participants clicked on a room map that depicted each of the rooms. Then they clicked on a seal map which represented each of the nine seals in that room. Clicking the icons on the seal map revealed the seal-container pair at that location. This mimicked the process of an inspector walking to different rooms in a facility and then finding the correct seals within each room. In half of the inspection conditions, the seal map updated throughout the inspection, indicating which seals had already been checked off of the inspector’s list. Our goal was to test the impact of dynamic updating on the participants’ ability to track their progress and find the remaining seals. With new and emerging technologies, it may be possible to provide safeguards inspectors with progress tracking or other dynamic updates while they are in the field, but it is not yet known if this will benefit their performance. Experiment 3 represents an initial test in this area.

4.1 Method

Participants. Eighteen participants were recruited from the employee population of Sandia National Laboratories and were compensated for their time. Two participants were later excluded from the analysis due to dropped eye tracking data for one of the experimental blocks. The remaining 16 participants (four female) had an average age of 34 years. Two of the participants held a high school degree, one held an associate degree, five held a bachelor’s degree, seven held a master’s degree, and one held a PhD.

Materials and Procedure. The seal-container pairs were the same as the items used in Experiments 1 and 2. As before, the inspector’s list was presented in two columns on the left side of the computer screen. The right side of the screen had a representation of a container and a seal similar to the one in Experiment 2. However, instead of using “Previous” and “Next” buttons to move between seal/container pairs, the participants had to navigate between virtual “rooms” to view the seals in each one. Below the representation of the container, there was a grid of nine seals which we referred to as the seal map. Below the seal map was the room map, which consisted of a row of four light gray squares labeled Room A, Room B, Room C, and Room D. See Fig. 6 for an example.

When participants clicked on one of the rooms in the room map, a message saying “Walking to Room X” appeared on the screen for two seconds and the room that they had clicked on turned dark gray, indicating that they were “in” that room. Participants could then click on the seals in the seal map, which made the corresponding seal and container numbers appear on the screen.

This experiment also had six conditions, but the manipulations were different from the other two experiments. There were three variants of the list presentation. The first

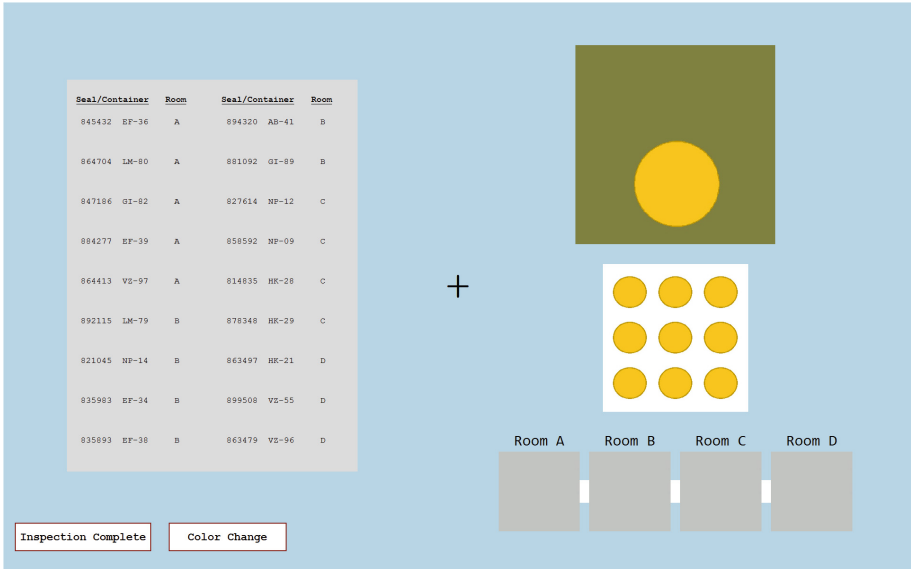


Fig. 6. Example of the screen layout used in Experiment 3. This example shows the room number condition, where the expected room was listed next to the seal/container pair in the inspector’s list. In the dynamic map conditions, the yellow seals in the 3 × 3 “seal map” turned gray as participants checked off the corresponding seals from their list. (Color figure online)

variant corresponded to the facility order condition in the prior experiments. The seal-container pairs were listed in the order in which participants would encounter them if they moved through the rooms and seals in the facility in order (row by row on the seal map) from Room A to Room D. The room in which the seal should be located was listed next to the seal-container pair in the inspector’s list. In the second variant of the lists, the seal-container pairs were listed in random order, but the room was listed with them. This was referred to as the room number condition. In the third variant, the seal-container pairs were listed in random order and no room information was provided. This was referred to as the random condition. In addition, there were two variants of the seal map, which could be static or dynamic. In the static condition, the seal map stayed the same throughout the inspection task. In the dynamic condition, the seal map updated based on which seals had been checked off of the inspector’s list. When seals were checked off, they were grayed out both on the list and on the map, so that participants could see at a glance which seals in each room had already been checked off. The list and map conditions were fully crossed, creating a 3 × 2 within-subjects design.

As in Experiments 1 and 2, there were eight items in the inspector’s list that were in the Match condition, two in the Wrong Container condition, two in the Missing condition, two in the Transpose condition, and two in the Transpose Match condition. These items were pseudorandomly assigned to the rooms such that there were never two of the same type of error in the same room. In addition, for the list conditions in which the room information was presented in the inspector’s list, there were two Match

items listed with the wrong room number. For example, a seal-container pair might be listed as being in Room A when it was really in Room C.

When participants made a decision about an item in the inspector's list, they clicked on it and six response choices appeared on the screen. The choices were the following: "Seal present, correct container, correct room," "Seal present, correct container, incorrect room," "Seal present, incorrect container, correct room," "Seal present, incorrect container, incorrect room," "Seal Missing," and "Other issue." Note that the fourth option, "Seal present, incorrect container, incorrect room" was not the correct response for any of the stimuli, but it was included to complete the set of possible response combinations.

Eye tracking data were collected for this experiment using the same procedure as Experiment 1.

4.2 Results

Accuracy. Across all blocks, the participants detected an average of 64% ($SD = 29\%$) of the background color changes in real time and 61% ($SD = 27\%$) of the color changes when asked to report the total number of changes that occurred during each inspection task. There was one participant who failed to report any of the color changes.

As in the other experiments, the participants performed near ceiling on all inspection conditions. The average percentage of seals categorized correctly ranged from 94% to 96%. A 3×2 within-subjects ANOVA showed that there were no significant main effects or interactions (all F s < 1).

Response Times. For each trial, the participants' response times were calculated as the time from trial onset to the time the participant clicked on one of the seals on the inspector's list. The average response times across all trials for each list presentation condition are shown in Fig. 7. A 3×2 repeated measures ANOVA showed that there was a significant main effect of list order ($F(2,75) = 9.00, p < 0.001$), and a significant main effect of map type ($F(1,75) = 9.08, p < 0.001$), but there was not a significant interaction between the two ($F(2,70) = 0.79$).

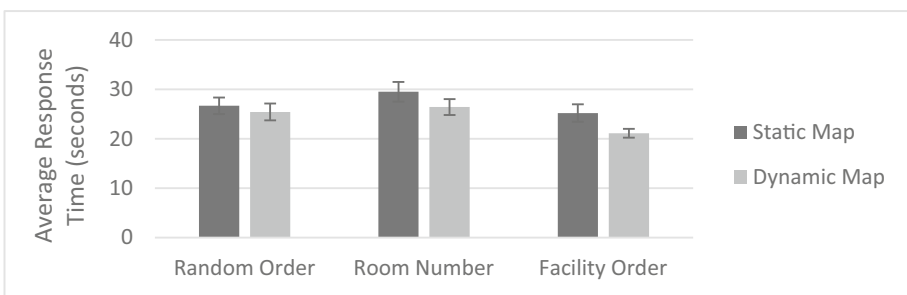


Fig. 7. Average response time for each trial for each of the six conditions in Experiment 3.

Post-hoc paired t-tests were used to assess the main effects. When participants had a dynamic map rather than a static map (collapsed across list order conditions), they had significantly faster response times ($t(15) = 3.18, p < 0.01$). Comparisons of the three list order conditions (collapsed across map type) showed that participants had significantly faster response times in the facility order condition than in the room number ($t(15) = 3.73, p < 0.01$) or random ($t(15) = 2.52, p < 0.03$) conditions. The participants had slightly faster response times in the random condition than in the room number condition, and the difference between the two was marginally significant ($t(15) = 2.08, p = 0.05$).

Search Strategy. All of the participants indicated that they searched by starting from one of the rooms and then searching the list for all of the seals in that room before moving on to the next room. An analysis of the number of times participants switched rooms indicated that there was not much variability across conditions, with the average number of switches per condition ranging from 6.1 to 7.4.

Seven of the participants mentioned the dynamic maps (or lack thereof) when describing their search strategies. Two participants wrote that they used the mouse to keep track of their place in the room when the maps did not update. Five of the participants mentioned that they double checked all of the seals in each room when the map did not update. For the dynamic maps, three participants wrote that they did a second pass through each room, checking only the seals that were not already greyed out. An analysis of the number of clicks on the seal map in each condition indicated that participants made many fewer clicks in the conditions with dynamic maps. These data are shown in Fig. 8. A 3×2 repeated measures ANOVA showed that there was a significant main effect of map type ($F(1,75) = 50.10, p < 0.001$), but not a significant main effect of list order ($F(2,75) = 0.51$), nor was there a significant interaction ($F(2,75) = 0.89$).

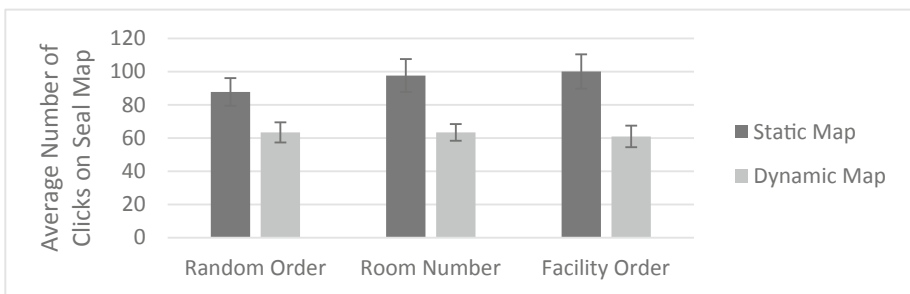


Fig. 8. Average number of seal views, as indicated by clicks on the seal map, for each of the six conditions in Experiment 3.

Eye Tracking Results. The ROIs in Experiment 3 consisted of each seal-container pair on the inspector's list, each room number on the inspector's list (when present), the container number and seal number in the "facility," the seal map, and the room map. The average proportion of fixations to each ROI was calculated for each

participant in each of the inspection conditions. This analysis revealed that the participants rarely fixated on the room numbers in their list, even though this information would have helped them to narrow their visual search of the list. The average percentage of fixations to the room numbers ranged from a low of 3% ($SD = 3\%$) to a high of 4% ($SD = 2\%$). The proportion of fixations to the seal map was also quite similar across all conditions, ranging from a low of 19% ($SD = 7\%$) to a high of 20% ($SD = 7\%$).

The average number of fixations per trial (defined for the eye tracking analysis as the time between participants' clicks on the screen, whether those clicks were on the list or on the maps) is shown in Fig. 9. A 3×2 repeated measured ANOVA showed that there was a significant main effect of list order ($F(2,75) = 11.25, p < 0.001$) and a significant main effect of map type ($F(1,75) = 7.90, p < 0.01$), but not a significant interaction between the two ($F(1,75) = 0.72$). Post-hoc paired t-tests showed that the dynamic map conditions had more fixations per trial in ($t(15) = 2.95, p < 0.01$), while the facility order condition had fewer than the other order conditions (both t s $> 2.87, p$ s < 0.02). The room number and random conditions did not differ significantly from one another ($t(15) = 1.89, p = 0.08$).

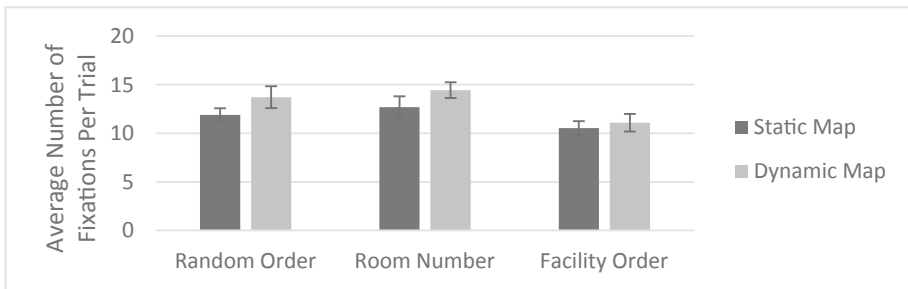


Fig. 9. Average number of fixations per trial for each of the six conditions in Experiment 3.

An analysis of which regions were fixated on each trial revealed a notable difference between the conditions with static and dynamic maps, as shown in Fig. 10. For the conditions with static maps, participants devoted approximately the same proportion of fixations to the item (the seal and container number) in the “facility” as to the inspector’s list. In contrast, for the conditions with dynamic maps, the participants devoted a higher proportion of fixations to the inspector’s list than to the item.

Paired t-tests showed that when there was a static map, there were not significant differences between the proportion of fixations to the item and the proportion of fixations to the list for the random or room number conditions (both t s < 1), although the difference approached significance for the facility order condition ($t(16) = 2.02, p = 0.06$). When participants had a dynamic map, there was a significantly higher proportion of fixations to the list for all list order conditions (all t s $> 5.51, p$ s < 0.001).

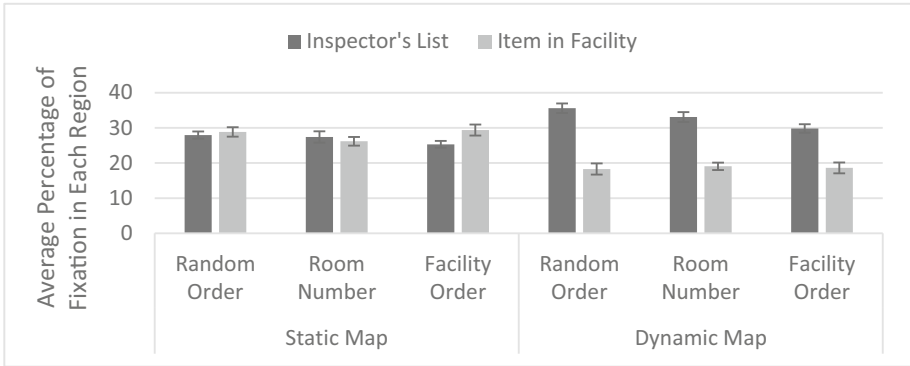


Fig. 10. Average percentage of fixations in each region per trial in Experiment 3.

4.3 Discussion

The results of Experiment 3 mirrored those of Experiment 2 in the sense that the list-to-item comparison led participants to search by looking at the item and then searching for a match in the list. Even though half of the items did not appear in the list at all, the participants universally used this search strategy rather than selecting an entry on the list and searching for it in the “rooms.” The participants generally did not take advantage of the room number information, which could have constrained their visual search of the list (or their search of the rooms, had any of them chosen to start their search from the list). This also mirrored the results of Experiment 2, where participants did not benefit from the presence of color coding in the list as an indicator of a seal’s location. We had predicted that participants might find room numbers easier to use than an arbitrary color code, but the participants did not seem to use this constraint. It is possible that checking the room numbers as they scanned the list for a specific seal or container number could have disrupted their ability to maintain the seal or container number in working memory. In addition, the room numbers were inaccurate for two items on every list, so the inaccuracies may have led the participants to mistrust the room numbers in general.

In the end, the room number condition had somewhat slower response times than the random order condition. The room numbers did not provide any benefit to the participants, yet they had to take the additional step of checking to make sure that the room numbers were correct. In the random condition, where no room numbers were listed, participants had fewer pieces of information to track for each item, leading to somewhat faster performance.

Although the list order conditions did not have a large impact on the participants’ search process, the map conditions did. Participants navigated from one seal to another by clicking on a seal map to reveal each item in a “room.” When participants had a map that updated to track which seals they had already checked off of their list, they made many fewer clicks on the seal map, made more fixations between clicks, and devoted a higher proportion of their fixations to the inspector’s list. At the same time, they were significantly faster to check the seals off of their list. This indicates that their visual

search process was more efficient. They did not need to spend time re-checking seals that they had already inspected. Instead, they could narrow in on the seals that had not been checked off and spend their time looking for those seals in the list.

Overall, the condition with the shortest response times was the facility order condition with a dynamic map. This indicates that both list ordering and progress tracking can provide benefits and that they can complement one another. Although the time savings for this condition is relatively small at the trial level (5–10 s, relative to the other conditions), it has an important cumulative effect. In a search task with 18 items, this per-trial time savings allowed participants to complete the entire inspection 1–2 min faster, on average, than the other conditions. In a real-world inspection, where inspectors may need to check dozens or hundreds of seals, the time savings provided by these kinds of supports could be substantial.

5 General Discussion

The experiments described in this paper demonstrate that list presentation can impact speed and efficiency for list-based visual inspection tasks. When participants are comparing two lists, as in Experiment 1, they tend to work from the shorter list and can take advantage of visual cues, such as color coding, that constrain which portions of the longer list they need to search. When participants are comparing a list to a set of items that must be viewed sequentially, as in Experiments 2 and 3, they tend to begin from the items and then search for them in the list, even if this means extra search time devoted to items that are not included in the list. In this situation, participants are better able to take advantage of list organization cues such as numerical ordering, whereas encodings of spatial information are less useful. When searching for an item in the list, participants have already located that item, and using that location information to narrow their search of the list would place an additional burden on their working memory. As a result, the participants generally ignored the encodings of spatial information, whether color coding or room numbers, and searched all of the remaining entries in the list. In all of these experiments, the inspector's list was fairly short, containing only 18 items. It is possible that inspectors would be more likely to take advantage of spatial cues if the list were longer, making the visual search more onerous.

In Experiment 3 we provided an additional support to the participants by providing them with dynamic information about which seals they had already inspected. This progress tracking made participants more efficient in their search and led to a considerable reduction in the number of times they viewed the seals, in addition to a reduction in their overall inspection time relative to other conditions.

Both sets of findings have important implications for the international nuclear safeguards domain, as well as other domains that involve list-based inspections. Although the IAEA inspectors do not have control over the materials provided to them by facility operators, they could change the formatting of their own materials to provide better support for their inspection process. The inspectors conduct both list-to-list and list-to-item checks, and the results of these experiments indicate that they are likely to use different search processes for these two types of tasks. Formatting their materials

accordingly and using new technologies to allow them to track their progress, when possible, will enable faster, more efficient inspections.

References

1. A day in the life of a safeguards inspector. <https://www.iaea.org/newscenter/news/a-day-in-the-life-of-a-safeguards-inspector>. Accessed 28 Jan 2019
2. Gastelum, Z.N., Matzen, L.E., Smartt, H.A., Horak, K.E., Moyer, E.M.: Brain science and international nuclear safeguards: implications from cognitive science and human factors research on the provision and use of safeguards-relevant information in the field. *ESARDA Bull.* **54**, 62–69 (2017)
3. Gastelum, Z.N., Matzen, L.E., Smartt, H.A., Horak, K.E., Solodov, A.A., Haass, M.J.: Testing human performance in simulated in-field safeguards information environments. In: *Proceedings of the 58th Annual Meeting of the Institute of Nuclear Materials Management* (2017)
4. Wolfe, J.M.: Guided search 2.0 a revised model of visual search. *Psychon. Bull. Rev.* **1**(2), 202–238 (1994)
5. Bednall, E.S.: The effect of screen format on visual list search. *Ergonomics* **35**(4), 369–383 (1992)
6. Ojanpää, H., Näsänen, R., Kojo, I.: Eye movements in the visual search of word lists. *Vis. Res.* **42**(12), 1499–1512 (2002)
7. Gramopadhye, A.K., Drury, C.G., Sharit, J.: Feedback strategies for visual search in airframe structural inspection. *Int. J. Ind. Ergon.* **19**(5), 333–344 (1997)
8. Hall, R., et al.: An evaluation of a job-aiding tool for an inspection task. *Proc. Hum. Factors Ergon. Soc. Ann. Meet.* **47**(18), 1987–1990 (2003)
9. Megaw, E.D.: Factors affecting visual inspection accuracy. *Appl. Ergon.* **10**(1), 27–32 (1979)
10. Megaw, E.D., Richardson, J.: Eye movements and industrial inspection. *Appl. Ergon.* **10**(3), 145–154 (1979)
11. Gallwey, T.J., Drury, C.G.: Task complexity in visual inspection. *Hum. Factors* **28**(5), 595–606 (1986)
12. Taylor, W., Melloy, B., Dharwada, P., Gramopadhye, A., Toler, J.: The effects of static multiple sources of noise on the visual search component of human inspection. *Int. J. Ind. Ergon.* **34**(3), 195–207 (2004)
13. Gastelum, Z., Matzen, L., Stites, M., Smartt, H.: Cognitive science evaluation of safeguards inspector list comparison activities using human performance testing. In: *Proceedings of the 59th Annual Institute of Nuclear Materials Management (INMM), 59th Meeting, Baltimore, MD* (2018)
14. Carter, R.C.: Visual search with color. *J. Exp. Psychol. Hum. Percept. Perform.* **8**(1), 127 (1982)
15. Green, B.F., Anderson, L.K.: Color coding in a visual search task. *J. Exp. Psychol.* **51**(1), 19 (1956)
16. Smith, S.L.: Color coding and visual search. *J. Exp. Psychol.* **64**(5), 434 (1962)
17. Williams, L.G.: The effect of target specification on objects fixated during visual search. *Percept. Psychophys.* **1**(5), 315–318 (1966)