SAND2018-6079C

Cognitive Science Evaluation of Safeguards Inspector List Comparison Activities using Human Performance Testing

Zoe N. Gastelum, Laura E. Matzen, Mallory C. Stites, Heidi A. Smartt

Sandia National Laboratories

Albuquerque, NM, USA

Abstract

While conducting in-field safeguards verification activities, international nuclear safeguards inspectors are confronted with information-rich environments in which they are expected to complete their tasks while maintaining high levels of situational awareness. In this setting, situational awareness is crucial both for their personal safety in an industrial environment and to enable the detection of any potentially inconsistent activities that could indicate an anomaly. Many in-field activities require inspectors to compare an IAEA-constructed list to other lists or items within the facility, such as physical inventory items, seals, records of shipments and receipts, or other documentation. The format of these lists could have a substantial impact on the ease or difficulty of these inspection tasks. How, then, can the IAEA format these types of lists so that safeguards inspectors can complete their inspection tasks with the highest levels of accuracy, timeliness, and situational awareness? In 2017, our team identified this area as one in which research could contribute to both the field of cognitive science and to international nuclear safeguards. In this paper, we will describe our research questions posited for human performance testing related to visual inspection and change detection tasks. We will then describe the experimental design and methods including our six experimental conditions for list presentation. Then we describe how task performance results were used to compare the list presentation conditions. We will close with an analysis and conclusions from the experiment, and proposals for additional research.

Introduction

The so-called information age has emerged from the convergence of the global spread of the Internet, cloud computing, and the miniaturization of technologies, and has led to pervasive information generation, access, sharing, and analysis (think of smartphones, which offer ubiquitous sensing with geo-location, still and video cameras, audio recorders, and more with the ability to instantaneously upload to social sharing platforms). Within this expansive and expanding information environment, the International Atomic Energy Agency (IAEA) is accessing and analyzing myriad new sources of potentially relevant open source and geospatial information streams. In addition, the IAEA has responsibilities in an increasing number of nuclear facilities across the globe as civilian nuclear energy programs emerge in newly nuclear states, and more states bring Additional Protocol declarations into force.

Within this vast safeguards information environment, safeguards-relevant information must be prioritized and provided to safeguards inspectors and analysts in comprehensible and actionable ways. In 2016, our research team began assessing how to best provide information to safeguards inspectors working in the field to support their timely, accurate, and situationally aware performance. We focus our research on information to which the IAEA Department of Safeguards currently has access and utilizes. While some

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

of these research findings may also be applicable to safeguards analysts working at IAEA Headquarters, we focus our research on safeguards inspectors conducting in-field safeguards activities.

In 2016 and 2017, we conducted an extensive literature review of both the cognitive science and international nuclear safeguards domains to identify unanswered research questions regarding how inspectors are provided with, use, and interact with, information while performing in-field verification activities [1]. Resulting from our literature review, the research team prioritized human performance research in three key areas: visual inspection, wayfinding, and knowledge transfer. In this paper, we will describe the experimental design, data collection, and results of our initial human performance experiment for visual inspection tasks for safeguards, and describe additional research that we propose would allow us to make broad recommendations to impact safeguards inspectors working in the field.

Research questions

We designed our visual inspection experiment around the physical inventory verification (PIV) and book audit activities in which IAEA safeguards inspectors compare a list from IAEA headquarters consisting of information from the state declaration (the "inspector list"), to an operator-provided list in the field ("facility list"). The purpose of the visual inspection experiment was to determine if the way in which the inspector list was presented could impact inspectors' timeliness, accuracy, and situational awareness during that activity. The metrics for each of these three criteria (timeliness, accuracy, and situational awareness) will be explained in more detail below.

We assumed for the experiment that the state- or operator-supplied information would be outside of the IAEA's control, and thus was held constant. Presentation of the "inspector" list, for this research, consisted of the visualization of the ordering and coloration of items in an electronically-presented list. Paper, 3-D, virtual or augmented reality, or other list visualizations were not tested. The test conditions used proxies for IAEA lists and inspectors, and will be described in additional detail below.

Method

Participants

Eighteen participants were recruited from the employee population of Sandia National Laboratories and were compensated for their time. Three of the participants were later excluded from the analysis due to failure to follow the task instructions (two participants) or poor eye tracking data (one participant). The remaining 15 participants (11 males and 4 females) had an average age of 37 years. The majority of the participants (12 of 15) held an advanced degree, including Master's degrees or PhDs.

Materials

The experimental materials consisted of six pairs of lists. For each pair, one list was designated as the inspector's list and one was designated the facility's list. Each list contained 36 pairs of seal numbers and container numbers, arranged in four columns. See Figure 1 for an example. The seal identification numbers were six-digit numerical strings. For each pair of inspector/facility lists, 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 attracted participants' attention or 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." There were ten possible letter pairs, each consisting of two unique letters (AB, CD, EF, HK, LM, NP, RS, TY and VZ). The same letter pairs were repeated multiple times in each list, but always with different numbers attached such that each container number was unique.

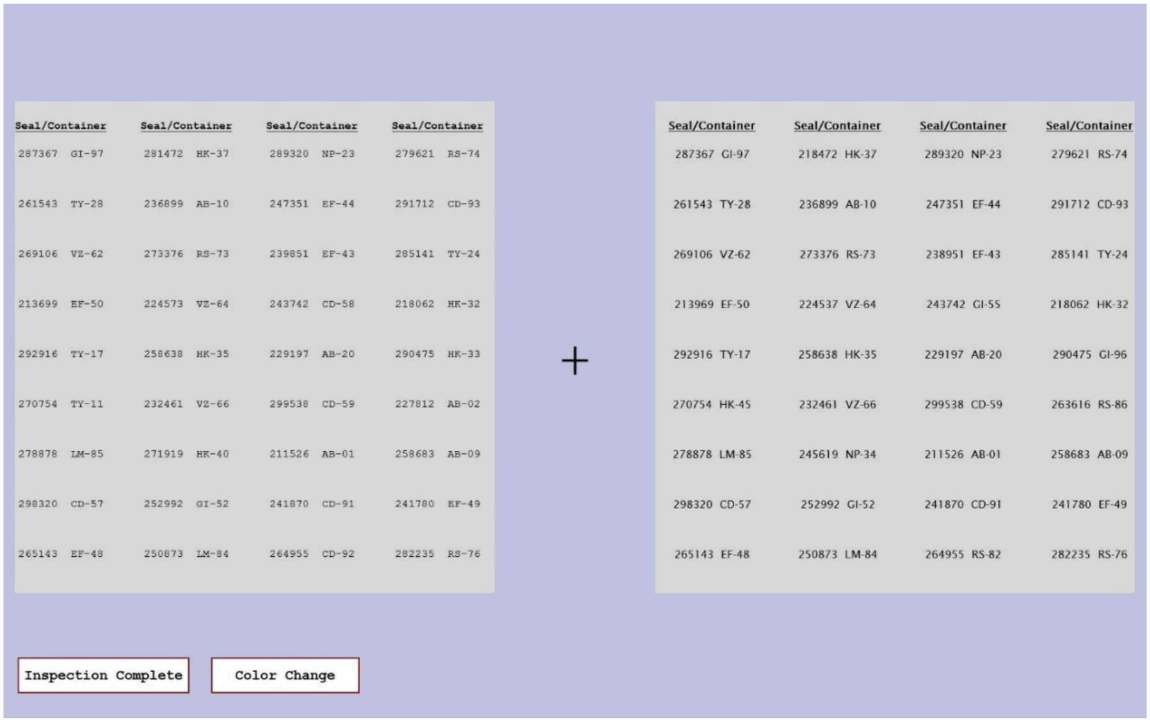

Figure 1. Example of the list pair stimuli, with the inspector's list on the left (in this case, randomly ordered with no coloration) and the facility's list on the right.

The items within each inspector's list were divided into five conditions. Of the 36 seal/container pairs on each inspector's list, 18 were matches to pairs on the facility's list. We refer to these as the "Match" items. Four of items on the inspector's list contained seals that were on the facility's list but paired with a different container number. For example, seal 270754 might be paired with container TY-11 on the inspector's list and with container HK-45 on the facility's list. We refer to these as the "Wrong Container" items. Two of the items on the inspector's list were entirely missing from the facility's list and are referred to as the "Missing" items. Four of the seal numbers on the inspector's list had a near-match on the facility's list, but with two numbers transposed. For example, the seal/container pair on the inspector's list might be "239851 EF-43" while the corresponding item on the facility's list was "238951 EF-43". We refer to these as the "Transposed" items. Finally, eight of the items on the inspector's list had both a match on the facility's list and a transposed version of the seal that appeared on both lists. For example, seal "261543" and seal "265143" both appeared on the inspector's list and both had a match on the facility's list. We refer to these as the "Transposed Match" items. These were included to increase the difficulty of the matching task. For the Transposed and Transposed Match items, there were four different types of transpositions: the 2nd and 3rd digit of the seal number, 3rd and 4th digit, 4th and 5th digit, or 5th and 6th digit. In half of each of these types of transpositions, the numbers were visually similar to one another (such as 6 and 9) and in half they were visually different (such as 1 and 2). The subtypes of transpositions were evenly distributed across all six lists.

The seal/container pairs in the facility's list were always presented in a random order. The order of the seal/container pairs on the inspector's list was manipulated across the six blocks. There were three ordering conditions for the seals: random order, 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, the items on the inspector's list were color-coded according to where the corresponding item appeared on the facility's list, where each column of the facility's list was assigned a color. An example of the random order, color coded condition is shown in Figure 2. Items on the inspector's list corresponding to those in the first column on the facility's list were colored red, and items corresponding to the second, third and fourth columns were colored teal, blue, and purple, respectively. The ordering and color-coding conditions were fully crossed, creating a 3x2 within-subjects design, meaning that all participants received all 6 list conditions.

Figure 2. Example of the random order, color coded condition

Task

The participants completed a list-to-list matching task that was divided into six blocks. Each block contained one of the six versions of the inspector's list, as described above. The inspector's list and the facility's list were presented side-by-side on the computer screen, with the inspector's list always presented on the left. Participants were instructed to check all the items on the inspector's list to see if they were on the facility's list. When participants made a decision about an item on the inspector's list, they clicked on it. Clicking on a seal/container pair brought up four response choices 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 and its corresponding container number were grayed out on the inspector's list, to help participants keep track of their progress. After checking off all of the items on the inspector's list, participants clicked a button to indicate that they had completed their inspection task.

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 purple and blue backgrounds are shown in Figure 1 and Figure 2, respectively). The changes were linked to particular 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. Participants were instructed to click on a button labeled "Color Change" as soon as they noticed a change in the background color. At the end of each block, they were also asked to report how many times the color had changed during that inspection task. The color change detection task was included as a measure of the participants' situational awareness during the inspection task. We hypothesized that participants would be more likely to notice the color changes when the inspection task was easier.

Procedure

After giving their informed consent, participants were seated in a dimly lit, sound attenuating booth. They were seated so that their eyes were 80 cm from the computer monitor. Participants completed a practice session that explained the task, including the background color changes and all of the possible configurations of the inspector's list. Then they completed a simple version of the inspection task in which they had to check off a list of four seals. After participants had completed the practice block and indicated that they understood the task, 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 of the experiment.

The participants completed the six blocks of the experiment in a random order. Each block began with a description of how the inspector's list would be organized. Each trial began with a fixation cross that was presented in the center of the screen for 1.5 seconds. Then the lists appeared on the screen and remained there until the participant clicked on a seal on the inspector's list. Clicking on a seal added the four response choices to the screen. After the participant clicked on one of the response choices, the next trial began. Clicking on the "Color Change" button also initiated a new trial. Once participants had checked off all of the seals on the inspector's list, they clicked the "Inspection Complete" button. They were then asked to indicate how many times the background color changed during the inspection task. Their choices ranged from zero to four, and they clicked on the number corresponding to their answer. Finally, participants were asked to give a brief description of the strategy that they used during the inspection task.

Results

Behavioral Data

The average reaction time and accuracy scores for the six different list organization conditions are shown in **Error! Reference source not found.**. The participants' overall accuracy was near ceiling for all conditions, while the reaction time results revealed much larger differences between conditions.

Accuracy

For the overall accuracy scores, a 3x2 within-subjects Analysis of Variance (ANOVA) was used to assess the differences across conditions. There was not a significant main effect of list order ($F(2, 70) = 0.56$) or of color coding ($F(1, 70) = 3.14$, $p = 0.08$). This is likely due to the near-ceiling performance for most of the stimulus conditions. The only stimulus type for which participants had lower accuracy scores was for the Transposed items. A 3x2 within-subjects ANOVA conducted for only the Transposed items showed that there was a significant main effect of list order ($F(2, 70) = 6.40$, $p < 0.01$), but no significant effect of color coding ($F(1, 70) = 0.71$) for these items. Post-hoc t-tests showed that for the conditions with color coding, participants had significantly lower accuracy for the Transposed items in the facility order condition when compared to the random order condition ($t(14) = 2.22$, $p < 0.03$). For the conditions without color coding, the average accuracy for the Transposed items was significantly lower in the facility order condition than in the numerical order ($t(14) = 2.09$, $p < 0.03$) or the random order ($t(14) = 2.81$, $p < 0.01$) conditions.

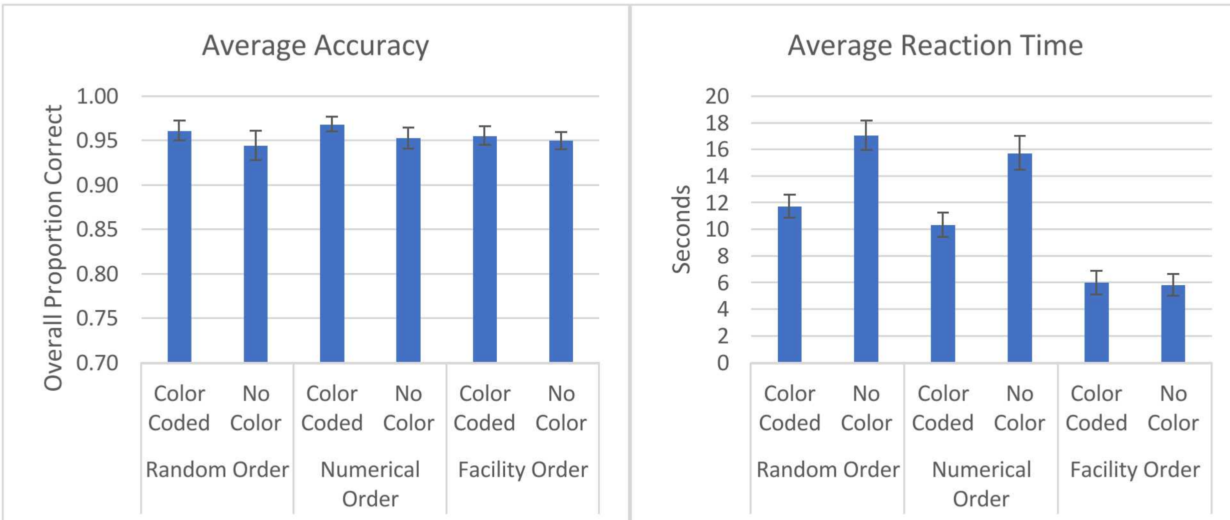

Figure 3. Average reaction time and accuracy scores for the six list conditions

Reaction Time

For the average reaction times, there was a significant main effect of list order ($F(2, 70) = 159.02$, $p < 0.001$), a significant main effect of color coding ($F(1, 70) = 70.74$, $p < 0.001$), and a significant interaction between the two ($F(2, 70) = 19.52$, $p < 0.001$). Post-hoc paired t-tests showed that participants responded significantly faster when there was color coding in the random order ($t(14) = 5.42$, $p < 0.001$) and numerical order conditions ($t(14) = 6.70$, $p < 0.001$). Color coding did not have a significant effect in the facility order condition ($t(14) = 0.56$). For the conditions with color coding, participants were significantly faster on the facility order condition than on the numerical order condition ($t(14) = 11.99$, $p < 0.001$) and significantly faster on the numerical order condition than on the random order condition ($t(14) = 2.41$, $p < 0.02$). For the conditions without color coding, participants were significantly faster on the facility order condition than on either of the other ordering conditions (ts > 10.79, ps < 0.001), and there was not a significant difference between the random and numerical order conditions ($t(14) = 1.42$, $p = 0.09$).

Search Strategy

The order in which participants checked off the seals on the inspector's list was analyzed to assess the strategy that each participant used to complete the inspection task. This analysis revealed three general patterns: checking off one row or column at a time, checking off one color group at a time, or using numerical ordering to locate seals. There were three cases in which a participant checked off the seals in a seemingly random order, in which case their order was classified as "other." Each list presentation condition supported different response patterns, so the presentation condition had a strong influence on the order of the participants' responses.

When the inspector's list was presented in a random order with no color coding, all but one participant used a systematic strategy in which they moved along one row or column at a time (the remaining participant checked off seals in a random order). When color coding was added to the random list order, four of the participants checked off one color at a time, ten moved along the rows or columns, and one used a random order. When the inspector's list was presented in numerical order, without color coding, four of the participants started their search from the facility's list and then used the numerical ordering of the inspector's list to guide their search. The other 11 participants moved along the rows or columns of the inspector's list and did not use the numerical ordering to their advantage. When the inspector's list had both numerical order and color coding, six of the participants checked off one color group at a time and nine moved along the rows or columns. None of the participants took advantage of the numerical ordering in this condition. When the inspector's list was in the same order as the facility's list, participants always moved along the rows or columns, regardless of whether there was color coding. There was only one exception, a participant who checked off the non-match seals first, then went back through the list to check off all the matching seals. The presence of color coding, which corresponded to the columns in these conditions, made the participants more likely search by column than by row. No participants moved row-by-row when the columns were color coded, but five participants moved row-by-row when they were not.

T-tests were used to compare the participants' average reaction times for different seal checking patterns in conditions where there was variability among the participants. For the numerical order condition with no color coding, participants were significantly faster when they took advantage of the numerical order than when they did not ($t(13) = 5.13$, $p < 0.001$). For the numerical order and random order conditions with color coding, there were no significant differences in reaction time between people who checked off items by color and people who checked off items by row or column (all ts < 0.97, all ps > 0.19).

Color Change Detection

During each block, the background color changed either two or three times. Participants were instructed to report the changes as soon as they noticed them. At the end of each block, they were asked to report the total number of color changes in that block. The average results are shown in Figure 4. Participants missed many of the color changes and there was a wide range of variability in their accuracy. 3x2 ANOVAs showed that there were no significant differences in color change detection across conditions (all Fs < 1.59).

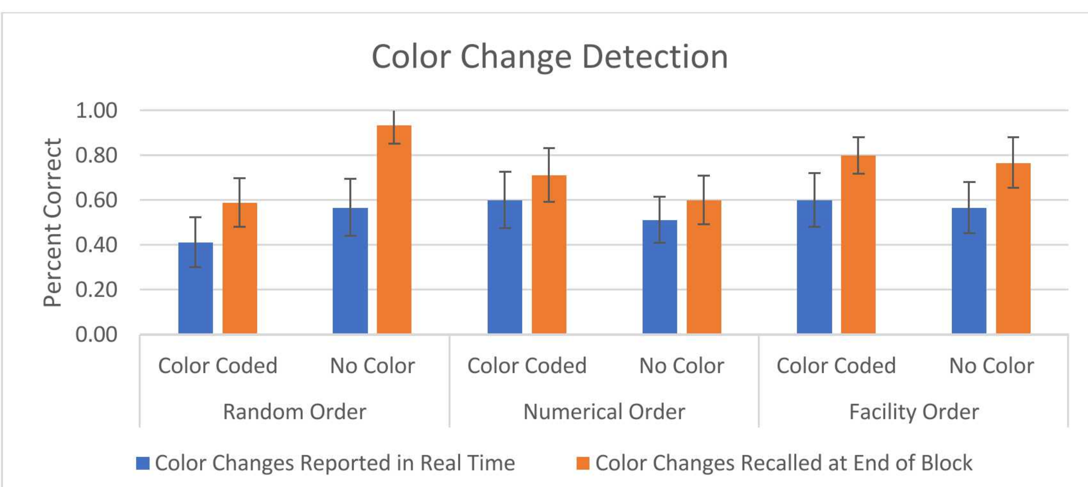

Figure 4. Average Accuracy for Detection of Color Change by List Condition

Eye Tracking Data

The average number of fixations per trial is shown in Figure 5. A 3x2 within-subjects ANOVA showed that there was a significant main effect of order ($F(2, 70) = 56.11$, $p < 0.001$), a significant main effect of color coding ($F(1, 70) = 34.07$, $p < 0.001$) and a significant interaction between the two ($F(2, 70) = 11.70$, $p < 0.001$). Post-hoc paired t-tests showed that participants had significantly fewer fixations per trial when there was color coding than when there was not for the random order ($t(14) = 5.02$, $p < 0.001$) and numerical order conditions ($t(14) = 3.22$, $p < 0.01$), but not for the facility order condition ($t(14) = 1.65$, $p = 0.06$). When comparing across the ordering conditions, participants had significantly fewer fixations for the facility order conditions than for the random and numerical order conditions, both with color coding (all ts > 5.07, all ps < 0.001) and without (all ts > 6.43, all ps < 0.001). There was not a significant difference in the number of fixations between the random order condition and the numerical order condition for either color coding condition (all ts < 1.29, all ps > 0.10).

To assess the efficiency of the participants' visual search, we calculated the average number of regions viewed per trial for each condition. This indicates how many seal/container pairs the participants looked at before finding the item that they were looking for or otherwise terminating their visual search. The results are shown in Figure 5. Note that there was a very wide range, with participants viewing an average of 10 seals per trial in the easiest conditions and over 60 in the most difficult condition. A 3x2 within-subjects ANOVA showed that there was a significant main effect of order ($F(2, 70) = 20.56$, $p < 0.001$), a significant main effect of color coding ($F(1, 70) = 18.73$, $p < 0.001$) and a significant interaction between the two ($F(2, 70) = 4.86$, $p < 0.02$). Post-hoc paired t-tests showed that participants viewed significantly fewer list items per trial when there was color coding than when there was not for the random order ($t(14) = 5.38$, $p < 0.001$) and numerical order conditions ($t(14) = 4.28$, $p < 0.001$), but not for the facility order condition ($t(14) = 0.20$). When comparing across the ordering conditions, participants showed the same pattern across both color coded conditions, viewing significantly fewer seals in the facility order conditions than in the numerical order conditions (all ts > 4.98, all ps < 0.001) and significantly fewer seals in the numerical order conditions than in the random order conditions (all ts > 1.79, all ps < 0.05). For a visual example of how the color coded condition changes participants' search strategies, see Figure 6. The color coding allowed participants to constrain their visual search to only one column of the facility list, making the search more efficient.

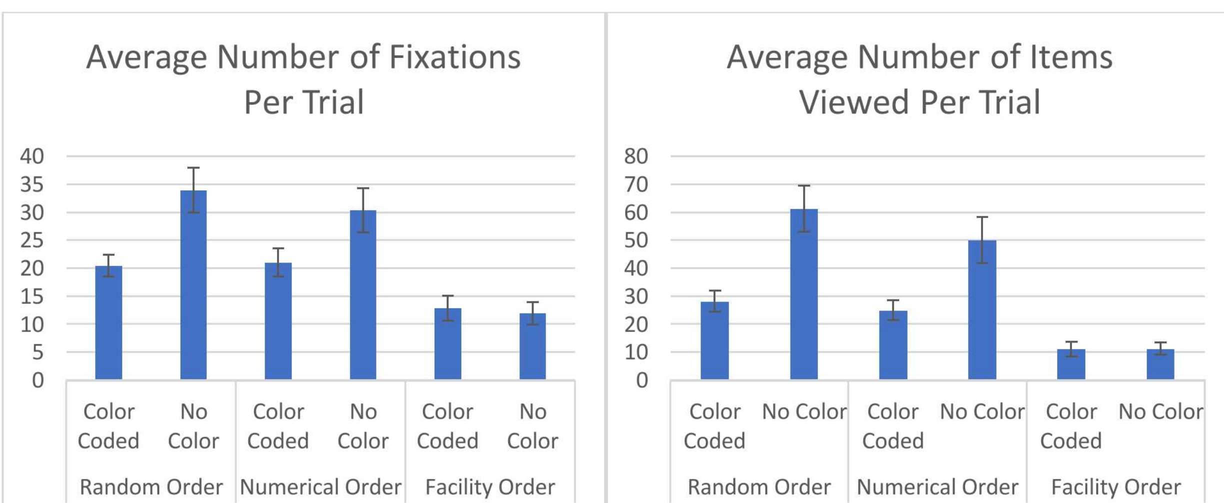

Figure 5. Average number of fixations per trial and average number of seal/container pairs viewed per trial for each experimental condition

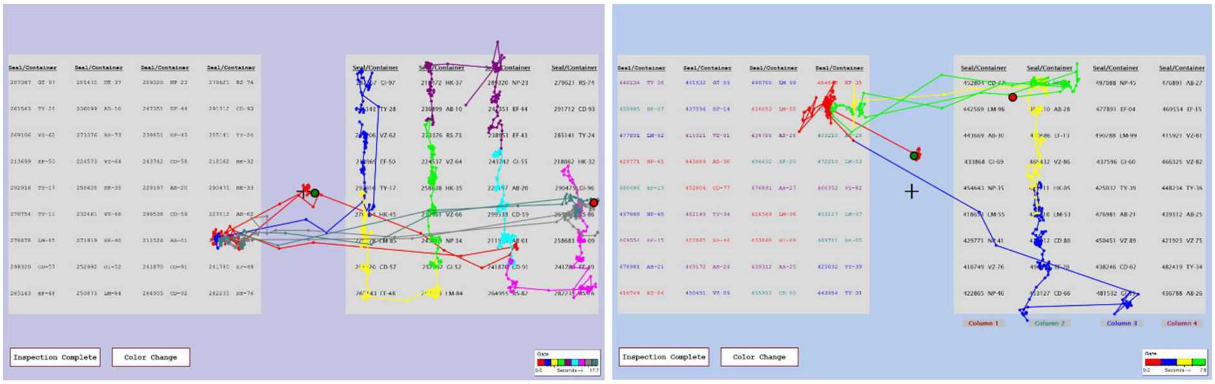

Figure 6. Example of how a color coded list condition impacts search strategy. The screenshot on the left shows a no coloration condition in which the participant searches up and down the columns on the facility's list. In the screenshot on the right, the participant used the color coded list condition to constrain eye movements within the approximate location of the searched-for item.

We also analyzed the patterns of eye movements for the trials where the background color changed. This analysis revealed a clear difference between the trials where participants noticed the color change and the trials where they did not. On trials where the participants noticed the change right away, they spent a large proportion of the trial looking at the background, with an average of 59% of the gaze points falling in the background region. When participants missed the color change, only 10% of the gaze points were in the background region. There were also cases where participants indicated that there was a color change when there was none. In these cases, an average of 64% of the gaze points fell in the background region, indicating that participants were studying the background color.

Discussion

The results of this experiment indicate that the ordering of an inspector's list does not have a substantial impact on the inspector's overall accuracy, but it has a major impact on how long it takes to complete a list-to-list comparison. When the two lists that the participants were comparing were in the same order, the participants completed the task fastest, had lowest average number of fixations, and viewed fewer seals on each trial, indicating a fast and efficient search process. However, one disadvantage of this rapid search process was that participants were less likely to notice transposed numbers in the seals. Although they had high accuracy overall, they missed some of the subtle differences between the two lists.

When the inspector's list was in numerical order, few people adapted their search strategies to take advantage of the list's structure. Participants generally started by selecting a seal on the inspector's list, searching for a matching seal on the facility's list, and then checking off that item on the inspector's list. Utilizing the numerical ordering required a different approach, namely selecting a seal on the facility's list and then using the numerical ordering of the inspector's list to quickly zero in on where that seal number should be. However, only four of the 15 participants altered their search strategy to look at the facility's list first in a no color condition. The participants who took advantage of the numerical ordering were significantly faster than those who did not. However, because so few participants took advantage of the order, the average reaction times for the numerical lists were very similar to the random order lists. If the numerical ordering approach were implemented in the field, it may be beneficial to train inspectors on the optimal search strategy to use in conjunction with this type of list structure.

In the numerical order, color coding condition, participants had the option of using either numerical order or color coding to guide their search. In that condition, none of the participants used the numerical order, even if they had used the numerical order to help them when there was not color coding available. The participants were substantially faster when relying on the color coding rather than the numerical order, indicating that the color coding was more effective in streamlining their visual search.

In the facility order condition, the color coding and the list order provided redundant information, so adding the color coding did not provide any additional benefit. However, the color coding had a substantial impact on the participants' speed for the other list order conditions. In the numerical order condition, participants were more likely to rely on the color cues than on the list order. Although participants were split between those who checked off items in color groups and those who checked of items by row or column, both strategies led to significantly faster performance when there was color coding compared to when there was not. This indicates that participants were using the color cues to guide their eye movements even when they did not change the order in which they moved through the inspector's list. This was also reflected in the lower numbers of fixations per trial and the lower number of seals viewed per trial in the color coding conditions. The color cues allowed participants to constrain their visual search, making it faster and more efficient.

Future Work

The visual comparison of two full lists was meant to simulate a physical inventory verification, or PIV, activity in which an inspector verifies all facility data. To understand a more complete scope of how inspectors use lists in the field, we plan to deploy two additional experimental conditions:

1. Partial list to full list. In this experiment, the inspector's list would represent a random subset of the facility's list, as with an interim inventory inspection or other random sampling done as part of routine safeguards inspections. This experiment is prompted by the idea that an individual's search strategy, and thus use of visual presentation of the list, may change based on whether they expect to verify the complete facility's list, or a random selection from that list. A partial list condition will make some search strategies observed in the full list comparison activity much less effective, and is thus expected to change the potential impact of list visualizations. It will also offer the research team the opportunity to more closely observe the impact of a "noisy" environment in which many of the state/operator list items will not be verified.
2. Partial list to a single item. In this experiment, the inspector's list would contain a sub-set of the items presented. This experiment represents an interim inventory verification or other random sampling done as part of a routine inspection that could apply to seal, container, or other item verification activities. Similar to the partial list to full list experiment, this experiment would have an additional "noise" factor because only a subset of the items presented to the inspector will appear on the inspector list. To accompany the "single item" experiment, we plan to implement an interactive map function that will further emphasize the facility location information. The map will allow participants to navigate between "rooms," each of which will have their own sub-set of seal/container pairs for verification. This should allow us to further examine search strategies based on the facility order condition.

References

[1] Zoe N. Gastelum, Laura E. Matzen, Heidi A. Smartt, Karl E. Horak, Alexander A. Solodov, Eric M. Moyer, "Testing Human Performance in Simulated In-Field Safeguards Information Environments." In Proceedings of the 58th Institute of Nuclear Materials Management Annual Meeting. Indian Wells, CA, USA, July 2017.