SAND2019-6871C

Human Performance Testing on Observation Capture Methods for International Nuclear Safeguards Inspections: Transferring Knowledge from the Field to Headquarters and Back

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

Sandia National Laboratories

Albuquerque, NM, USA

ABSTRACT

Knowledge transfer refers to the sharing of information or experience across teams or organizations. Within the field of cognitive science, knowledge transfer has been studied extensively in the contexts of long-term institutional knowledge preservation and short-term transfer of information between teams in shift work environments, where lapses in knowledge transfer can result in high-consequence accidents or errors. Both types of knowledge transfer are relevant to international nuclear safeguards, yet safeguards work also presents unique constraints that have not yet been studied. For example, studies of shift work examine real-time, face-to-face handoffs of information between individuals or teams. In the international nuclear safeguards domain, it can be days or weeks between the time an inspection team records their observations in the field and when they brief colleagues and management at headquarters, followed by additional weeks or months before that information is acted upon in the field. Further, while inspectors are restricted in the information that can leave a nuclear facility, they are generally permitted to take hand-written notes and, in some cases, can take (or request an operator to take on their behalf) digital photographs to be kept securely onsite or transferred back to headquarters. Some guidance is given to inspectors on how to record observations and utilize digital photography, but it is largely anecdotal. In response, the research team developed and conducted human performance studies in a safeguards-relevant environment to compare the efficacy of different mechanisms for recording observations in supporting the detection of different types of changes at a later time. In this paper we will describe our experimental design, results, and recommendations for additional research.

INTRODUCTION

Nuclear safeguards inspectors for the International Atomic Energy Agency (IAEA) visit nuclear sites around the world to verify that states are conducting their nuclear activities according to their respective safeguards agreements and have not diverted nuclear material, misused nuclear facilities, or otherwise developed undeclared nuclear activities. As part of the safeguards inspections and visits, safeguards inspectors make notes and observations, including the use of digital photography where permitted, about nuclear facilities to be used by themselves or their colleagues on future visits to understand any changes in activities at the facility. Recording notes and observations for international safeguards is critical for preserving important knowledge learned in the field. Inspectors' notes are used to prepare briefings to management and State Evaluation Groups upon return to IAEA headquarters (which may be days to weeks after the visit), and to help prepare the next group of inspectors that will be visiting the site.

Inspectors receive guidance for both note-taking (including the note-making approach) and the use of digital photography (for example, making notes of locations where images are taken). Anecdotally, though, inspectors record notes and observations using the methods most comfortable to them.

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.

The practice of knowledge transfer – the sharing of knowledge, information, and experience between people within an organization (Argote and Ingram 2000) – has been well-studied in the cognitive science community, especially as it relates to shift workers such as those working in production (Bosua and Venkitachalam 2015), medical environments (Kerr 2002; Patterson et al. 2004), off-shore drilling (Lardner 1996), and nuclear power plants (Wilkinson and Lardner 2013). However, safeguards inspections pose a unique research opportunity compared to more traditional shift work environments due to the time lag (days to weeks) between the initial information collection and briefing, and then again (weeks to months) before the next on-site activity. This leads us to ask: *how well do different note-taking and observation recording techniques work to preserve knowledge over longer time spans than previously studied, both for one's self and for passing that knowledge to others?*

RESEARCH QUESTION

In this study, we elected to examine how inspectors document their observations in the field, and how those notes can be used later to identify changes in a facility. Specifically, we asked how well an inspector could identify changes to abstract drawings based on their method of note-taking, including an examination of the use of pencil and paper, digital cameras, a combination of written notes and digital photography, or no notes at all. One of the most novel aspects of the experiment was our use of a time delay, in which participants returned to the laboratory approximately two days after taking their notes to complete a memory task while using their original notes as a memory aide.

The three research questions for the work reported here were:

1) What is the most effective note-taking method for transfer of complex visual knowledge?
2) How does note-taking method impact one's confidence in assessments of changes?
3) Do some note-taking methods enable more efficient knowledge transfer, measured by time to complete the study?

METHODS

Participants

We do not have access to conduct human performance studies directly on IAEA safeguards inspectors. As a proxy, twenty-one participants were recruited from the staff of Sandia National Laboratories. One participant was excluded due to experimenter error administering the study. The final study population of twenty participants was comprised of seven females and 13 males, with a mean age of 44 years (range: 24-68). Participants provided their informed consent prior to participation, and all procedures were approved by Sandia's Human Studies Board.

Materials

Participants in the study saw two sets of images: study images and test images. Both sets of images contained computer-generated graphics of abstract industrial objects (e.g. widgets, gears, etc.) which were generated by the U.S. Intelligence Advanced Research Projects Activity (IARPA) Machine Intelligence from Cortical Networks (MICrONS) project, and generously shared with our research team.[1]

[1] Only those images which have been made available to MICrONS researchers are published, to ensure robustness of MICrONS evaluations.

Each set of images contained four study boards consisting of an eight-by-five matrix of images: 40 images per board, 160 total images. For the test boards, 20 images per board were modified: four depicted objects that were of a different material than the original, four objects were rotated to a new orientation, six objects were replaced with new objects, and six objects were moved to other locations on the board (see Figure 1). This resulted in 56 replacement images, for a total of 216 unique images viewed over the course of the experiment.

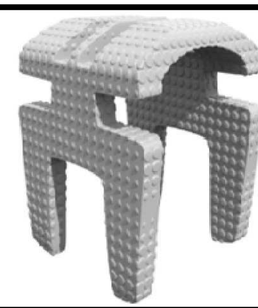
Type of change	**Before**			**After**		
Material						
Orientation						
Location	A	B	C	E	B	C
	D	E	F	D	A	F
	G	H	I	G	H	I
Replaced by different object	A	B	C	A	B	C
	D	E	F	D	X	F
	G	H	I	G	H	I

Figure 1 Example of each change Type. Image Credit (for material and orientation images): IARPA

Participants were supplied with all the note-taking supplies they needed to complete the procedure described below, including a pencil, pad of paper with clipboard, and digital camera.

Procedure

The experiment consisted of two sessions: a study session and a test session. The study session and the test session were conducted approximately two days apart. The study session began with an overview of both sessions.

During the study session, participants were presented with each of the four study boards in series, which they were instructed to observe and make notes for a later memory test on image layout and characteristics. The note-taking conditions (one per study board) were:

1. **Handwritten notes**, for which the participants were instructed: “You can take notes in this notebook to help yourself remember the layout of the images. You can use any combination of words or pictures.”
2. **Digital camera**, for which the participants were provided with a Nikon Coolpix camera similar to those used by IAEA safeguards inspectors for Complementary Access visits. Participants were instructed: “You can take photos of the board with the digital camera provided. The only restriction is that you can’t take a picture that includes the entire board, but you can take pictures of the individual items, or groups of items together.”
3. **Handwritten notes and digital camera**, for which participants were instructed: “You can take photos of the board with the digital camera provided (with the restriction that you can’t take a picture that includes the entire board). You can also take notes in the notebook to help yourself remember anything about the layout of the images, or your strategy while taking the pictures.”
4. **No aid**, for which participants were instructed: “For this board, you can only rely on your own memory; you cannot take notes in any way.”

Participants were assigned one of four note-taking conditions for each study board, so that all participants had all four conditions evenly distributed across the boards and evenly distributed across the order in which they completed an assigned condition. Participants had up to 12 minutes to review (and make notes, if applicable) each study board, but were not required to use the whole 12 minutes. Participant time for each study board was recorded in seconds.

Approximately two days after the study session, participants returned for the test session. Recall that the test boards contained 20 changes out of 40 total images, including material, orientation, board location, and images that were replaced with new, unstudied images. In the test session, participants were asked to recall the image boards they viewed during the study session, and record any differences using their notes from the study session as a guide. Participants were provided with an sheet of paper containing a grid of 40 squares that matched the layout of the test boards, with one square per item on the board. Each square contained two yes/no questions: 1) Did the object change? and 2) Are you sure? Participants recorded their assessments by circling Y or N. If participants recorded a change, they were asked to write a description of the change. They were shown examples of the four possible types of changes in a handout similar to Figure 1**Error! Reference source not found.**, and were allowed to refer to the handout throughout the test session.

Participants were provided with their original notes for each board. Test boards were viewed in the same order that they were studied. As with the study session, participants were allowed up to 12 minutes to complete their response sheet for each test board and could elect to finish early. Participant time on each board was recorded in seconds. Total experimental time, including instructions and participant questions, lasted no more than one hour per session.

RESULTS

For each item, participants scored a 1 if they correctly answered that an item had changed, and a 0 if they answered incorrectly. For items that had changed, participants scored a 1 if they correctly identified the type of change, or a 0.5 if they mistook a new item for a location change (since for that particular location, a location change can appear to be a new item if the participant doesn't observe the new location). For select cases in which an orientation change made an item appear so different that multiple participants mistook it for a new item, participants also scored 0.5. Participants scored a 0 if they reported any other incorrect change type. Confidence ("are you sure?") scores were also converted to numeric values, with 1 for "yes" and 0 for "no."

We used the *d'* statistic to show the participant's ability to discriminate targets (hits and correct rejections) from distractors (false alarms and misses). Higher *d'* values indicate that the participant more frequently responded "change" to changed items and rarely to unchanged targets. Lower *d'* values indicate the participants frequently responded "no change" to changed items and "change" to unchanged items.

<u>Effects of Note-taking Condition and Object Changes on Change Recognition</u>

Our analysis showed that there was a main effect of note-taking condition and change type on a participant's ability to correctly recognize a change. Participants were most able to identify changes in the *camera only* and *camera + notes* conditions (though there was not significant difference between the two conditions). Location changes and new items had the highest rates of correct assessments, followed by orientation changes, with material changes having the lowest overall scores. The *notes only* condition had much higher target discrimination rates than the *no aid* condition for location and object changes but performed approximately the same as the no aid condition on the material and orientation changes (see Figure 2).

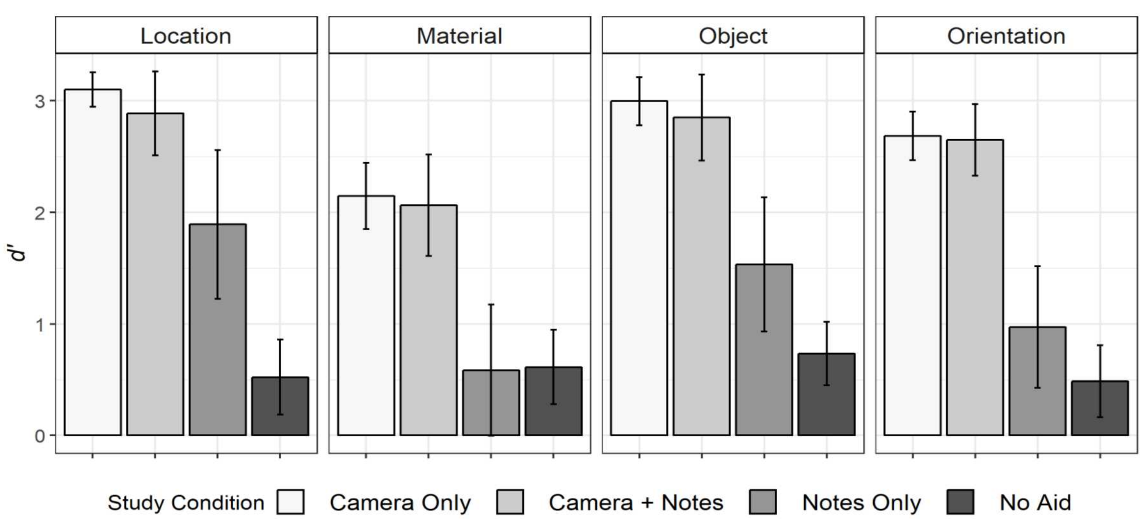


Figure 2 Bars show the d' score for each study condition and item change type. Errors bars represent 95% confidence intervals around the mean.

This suggests that hand-written notes alone were better than no aid in transferring broad information about the image matrix, like the identity and overall layout of items, but were less effective for transferring information about more subtle changes, like the item's material or orientation.

Effect of Note-taking Strategy on Change Recognition

For the experimental conditions in which participants took notes (*notes only* and *camera + notes* conditions), they were free to take whatever style of notes they preferred and were reminded in the directions that this could include text, drawings, or a combination of the two. Though we had not intended to study note-taking styles, we hypothesized after collecting the participant notes that participants' self-selected note-taking type may have impacted their ability to discriminate changes to items on the photo boards.

We binned the participant notes according to three general categories, including: photo layout descriptions (i.e. a diagram of the entire photo matrix), verbal descriptions of the objects, and hand-drawn pictures of the items. See Table 1 for the note types and average accuracy metric for each group. There was no significant difference by note type in the *camera + notes* condition (potentially because there was a wide variety of note-taking strategies employed), but we did observe a trend that participants who included multiple note-taking strategies (e.g. photo layout and a verbal description) performed better than those with only a single strategy.

Table 1 The table lists d' scores for each participant-selected note-taking strategy within the Camera + Notes and Notes Only Study Conditions.

Study Condition	Note-taking strategy	N	*d'*	Standard Deviation
Camera + Notes	Photo layout	8	2.76	0.56
	Photo layout + drawings	2	2.78	0.08
	Photo layout + verbal descriptions	1	3.12	NA
	Drawings + verbal	1	2.94	NA
	Drawings only	1	2.46	NA
	Verbal only	6	2.21	1.31
	None	1	2.81	NA
Notes Only	Drawings only	8	1.12	0.74
	Verbal only	3	-0.23	2.65
	Drawings + verbal	9	1.85	0.68

For the *notes only* condition, we binned the notes differently since the photo-layout was not as useful without the digital images. Instead, we described them as: drawings only, verbal only, or a combination of drawings and verbal notes. Here we saw a significant effect of note-taking strategy, in which the combined drawings and verbal had significantly better performance, followed by the drawings only. The verbal only notes had the lowest levels of performance.

This suggests that taking more elaborative notes that included both verbal and verbal descriptions led to better performance than using a single type of encoding (i.e., either drawings *or* verbal descriptions). The negative *d'* score for verbal-only notes likely indicates that participants were guessing.

Change Type Accuracy

When participants correctly answered that a change in a test item occurred, we asked whether their ability to identify the type of change differed across study condition. Recall that half of the items on each test board changed. Items that changed and were correctly identified as changed comprised 1129 observations, or 35% of the original dataset. (One participant was excluded from this analysis due to a failure to answer any change correctly in one condition).

Our analysis showed that note-taking condition had a significant effect on change type accuracy. The *camera* and *camera + notes* conditions performed comparably. Both were significantly better than the *no aid* condition, although only the *camera + notes* condition was significantly better than the *notes only* condition (see Figure 3). This pattern of findings suggested that the use of both hand-written and digital notes provided the greatest benefit in terms of transferring information about the type of change an item underwent (especially compared to notes alone), while the *camera only* and *notes only* conditions were not statistically different from each other.

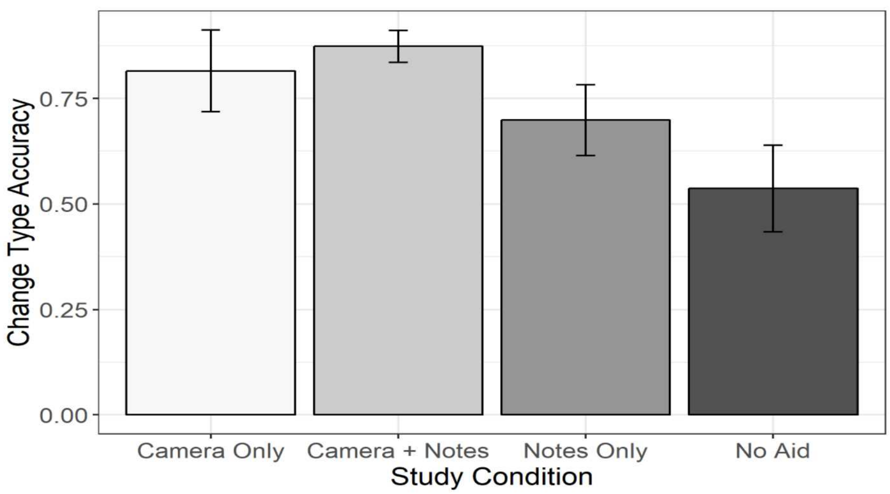


Figure 3 Change type accuracy by study condition. Error bars are 95% confidence intervals around the mean

Confidence

For each item they assessed, participants were asked if they were sure of their answer. Their response was converted into a numeric value (Sure=1, Not Sure=0) and averaged for each participant by study condition. We found that study condition had a significant effect and differed between all conditions except those that included a camera (*camera only*, *camera + notes*), which were statistically indistinguishable. This pattern was similar for the relationship between participant "sureness" and their *d' scores*, indicating that participants recognized that their ability to identify changes was best in the *camera only* and *camera + notes* conditions.

We also asked how confidence differed across accuracy types, that is for *hits* (correctly said "change" when a change occurred), *correct rejections* (correctly said "no change" when no change occurred), *false alarms* (incorrectly said "change" when no change occurred), and *misses* (incorrectly said "no change" when a change occurred). Confidence rates for hits were similar across all three note-taking conditions, although the *notes only* confidence was slightly lower than the two camera conditions. Confidence rates were substantially lower in the *no aid* condition for all accuracy types. Confidence for correct rejections dropped for the *notes only* condition, indicating that using a camera enabled participants to more confidently say when no change had occurred. However, confidence for misses dropped substantially for the *notes only* condition but remained high for the *camera only* and *camera + notes* conditions. This pattern suggests that using the camera as a study aid, with or without additional notes, may have over-inflated confidence levels, particularly when participants incorrectly reported that no change was present (see Figure 4).

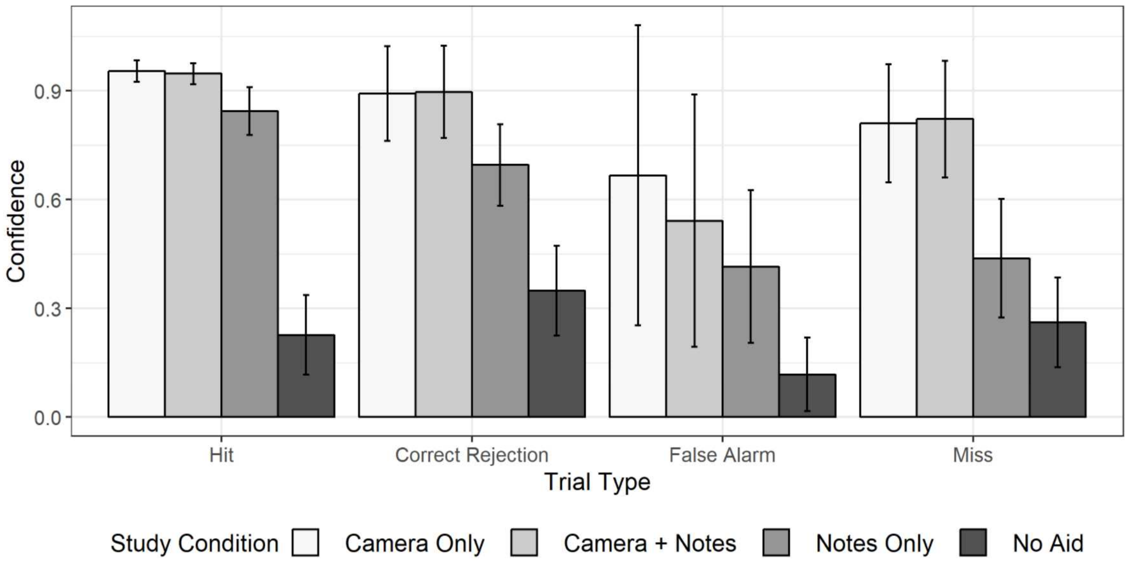


Figure 4 Mean confidence ratings across study conditions, split by item accuracy. Errors bars represent 95% confidence intervals around the mean.

<u>Response Time</u>

Response time was measured as the length of time participants took for each condition on both the study and test boards, measured in seconds. Study and test times for each condition were limited to 12 minutes (720 seconds).

Study time was longest for the *notes only* condition, followed by the *no aid* condition, the *camera + notes* condition, and the *camera only* condition. The *camera + notes* and the *no aid* conditions did not differ significantly. While the *camera + notes* condition took longer in the study session than the *camera only* condition, the only benefit to knowledge transfer was the numerically (but not statistically significant) ability to correctly identify an item's change type. See Figure 5 for study times.

Test times were longest for the *camera + notes* condition, followed by *camera only*, *notes only*, and then the *no aid* condition (see Figure 5), but the only statistically significant pair-wise comparisons were between the *camera + notes* and the *no aid* condition. A corresponding benefit to study time for the *camera only* condition was not observed at test, suggesting that while the *camera only* condition may have made the study condition easier, the difficulty of accessing and using the photos during the test condition negated the benefit.

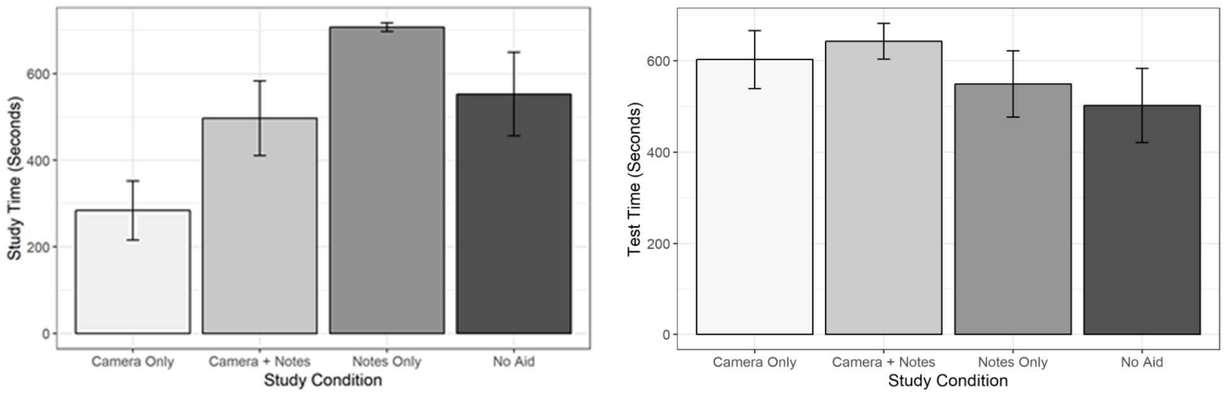


Figure 5 Average time across conditions for study (left) and test (right). Errors bars represent 95% confidence intervals around the mean.

DISCUSSION

Our results indicated that the *camera only* and *camera + notes* conditions produced the highest levels of knowledge transfer, indicated by higher *d'* scores which were statistically indistinguishable from each other across our analyses. The *notes* condition was better than *no aid* for transferring knowledge about image locations and replacements but performed equally for more subtle material and orientation changes.

When participants had access only to written notes, we observed that more elaborate notes written with both drawings and verbal descriptions had the highest performance relative to participants who used exclusively drawings or exclusively verbal description. Because this was a post-hoc observation and not equally counter-balanced across our participants, it could have been driven by the elaboration effect at encoding (Kobayashi 2005) or increased ease of recall at test (Kiewra 1989). However, it still suggests that when hand-written notes are the only available study material, multiple information recording strategies (i.e. both visual and verbal) may be most supportive for knowledge transfer.

When using the *camera + notes*, participants numerically had the highest performance. The weak evidence that *camera + notes* performed better than *camera only* could be explained by the study time, in which the *camera + notes* participants took more time to process and consolidate the information by making their own drawings or descriptions rather than mechanistically capturing all information exactly with the camera.

Further, when using camera information participants showed higher confidence overall. This included overconfidence in incorrect responses, especially when they missed a change. Finally, while using the *camera only* condition, participants had the shortest study time, but the benefit did not translate to the test time. Anecdotally, participants complained about interaction with the digital camera at test and appeared to struggle to use the device, indicating that benefits of using the camera only might be outweighed by the difficulty of using it at test.

RECOMMENDATIONS

Based on our findings, we recommend that, when allowed, inspectors should use digital cameras to capture photos at a site. Time permitting, hand-written notes should accompany the photos. When inspectors make hand-written notes, elaborative notes including multiple types of descriptions (i.e. both visual and verbal) should be captured. If using only handwritten notes, some changes like surface material and object orientation are more difficult to transfer, and extra attention should be paid to recording those details.

FUTURE RESEARCH

Two follow-on studies are being conducted. In the first, we invited back the 20 participants who were used in the analysis of the first study. They returned to the lab 6-9 months after their initial participation and completed the test session for a second time. They were given access to their original notes for the *camera*, *camera + notes*, and *notes only* conditions. For the *no aid* condition, participants were given notes that were taken by a different participant in the original experiment. We used four sets of notes, one of which was exclusively text-based, one of which was exclusively drawings, and two that contained a combination of drawings and text. In one of these sets of notes, the notes are laid out in a grid that matches the layout of the study board, while in the other, the notes are presented in a list. We will assess participants' performance using their own notes after a substantial delay, as well as their performance when using a different person's notes.

In the second follow-on study, new participants will view four test boards and attempt to fill out the response sheets for each board while relying exclusively on other people's notes. This study will use the same four sets of notes that were used in the first follow-up study. This study will allow us to assess the impact of these different note styles on performance in a within-subjects design.

ACKNOWLEDGEMENTS

This work was supported by Sandia National Laboratories' Laboratory Directed Research & Development (LDRD) funding. Thank you to IARPA for use of MiCRONS images as our stimuli, and to Aaron Jones for his experimental administration.

REFERENCES

Argote, L., Ingram, P.: Knowledge transfer: A basis for competitive advantage in firms. Organizational Behavior and Human Decision Processes (82), 150-169 (2000).

Bosua, R., Venkitachalam, K.: Fostering knowledge transfer and learning in shift work environments. Knowledge and Process Management 22, 22-33 (2015).

Kerr, M.: A qualitative study of shift handover practice and function from a socio-technical perspective. Journal of Advanced Nursing 37, 125-145 (2002).

Kiewra, K.: A review of note-taking: The encoding-storage paradigm and beyond. Educational Psychology Review 1(2), 147-172 (1989).

Kobayashi, K.: What limits the encoding effect of note-taking? A meta-analytic examination. Contemporary Educational Psychology 30, 242-262 (2005).

Lardner, R.; Effective Shift Handover: A Literature Review; Offshore Technology Report – Health and Safety Executive OTO (1996).

Patterson, E., Roth, E., Woods, D., Chow, R., Gomes, J.: Handoff strategies in settings with high consequences for failure: Lessons for health care operations. International Journal for Quality in Health Care 16(2), 125-132 (2004).

Wilkinson, J., Lardner, R.: Pass it on! Revisiting shift handover after Buncefield. Loss Prevention Bulletin 229, 25-32 (2013).