SAND2014-2122C

The Effect of Job Performance Aids on Quality Assurance

Erik Fosshage, Sandia National Laboratories / Purdue University

Job performance aids (JPAs) have been studied for many decades in a variety of disciplines and for many different types of tasks, yet this is the first known research experiment using JPAs in the quality assurance (QA) context. The objective was to assess whether a JPA has an effect on the performance of a QA observer in the concurrent dual verification construct using a simple checklist for a basic assembly task. Results show that the JPA has only a limited effect, however, there were 3 significant findings that may draw interest from a variety of practitioners. First, a novel testing methodology sensitive enough to measure the effects of a JPA on performance was created. Second, the discovery that there are different probabilities of detection for different types of error and their impact on the QA context may be the most far-reaching result. Third, these results highlight the limitations of concurrent dual verification as a control against defects.

INTRODUCTION

According to James Cantrell, the main engineer for the Skipper satellite, "It's always the simple stuff that kills you." Skipper failed one day into its mission because Russian scientists mistakenly connected the solar panels backwards ("Russians Miswire Satellite", 1996). This example reminds us that even simple errors can have large consequences, and though human error cannot be eliminated processes can be designed to prevent errors or minimize their impact. This study examines how a JPA may affect human performance in the verification activities that are often employed in systems that have high consequences for failure.

LITERATURE REVIEW

Probably the best-known JPA is the pilot's checklist, which traces to the initial test flight of the Boeing Model 299 aircraft on October 30, 1935. The Model 299 crashed after take-off but the investigation determined that no mechanical failures had occurred and that the aircraft crashed because the pilot forgot to unlock the elevator and rudder locks. The Model 299 was substantially more complex than previous aircraft and simply too much for one pilot to remember how to fly, so the approach at the time was to simplify this complexity with a pilot's checklist (Meilinger, 2004; Gawande, 2010). Over the next few decades research on job performance aids (JPAs) would be conducted in a variety of disciplines. Miller (1956), Newman (1957), and Miller (1953) studied how behavioral and psychological factors could be identified and then used to design training requirements for specific tasks, and it was during the latter study that Dr. Edgar Shriver coined the term "task analysis" to specify the need to identify stimulus elements, or behavioral cues, that indicate when a task is complete and then place these directly into training instructions (Shriver et al., 1982). Task analysis was widely adopted and matured by a variety of practitioners (e.g. Folley, 1961; Goff et al., 1969) and the JPA was found to be an effective tool at simplifying tasks for novice users that would normally require extensive training or complex information processing. These experiences led to a variety of specifications (e.g. Folley et al., 1971; Joyce et al., 1973a; Shriver, 1975) and handbooks (e.g. Joyce et al., 1973b; Smillie, 1985) to assist developers in creating effective JPAs for a wide variety of tasks in both military and commercial industries.

Meanwhile, specific industries were conducting their own studies on the limitations of human performance. McKenzie (1958) was an early proponent of having clear standards and instructions in order to improve the accuracy and consistency of inspectors. This followed with more than 3 decades of research from the industrial inspection community (summarized in detail by Wiener, 1975, and See, 2012), with some of it focused on vigilance theory (e.g. Mackie, 1964; Tsao, Drury, and Morawski, 1979) that stemmed from the pioneering work of Mackworth (1950). These were followed by researchers who studied underlying models of human information processing that are useful to JPA developers (Norman, 1981; Rasmussen, 1982; Reason, 1990). Following the Three Mile Island accident the nuclear power industry commissioned JPA research of its own (Clark, 1982; Shriver et al., 1982; Mattson, 1989), and in the years since other high consequence industries have followed suit by embracing JPAs and wider principles of error management, human factors, and quality assurance (Helmreich and Merritt, 2001; Haynes et al., 2009; White et al., 2010).

Sandia National Laboratories (SNL) designs and builds a variety of high consequence products, however its legacy is with nuclear weapons and as such is subject to the guidance and oversight by the Department of Energy (DOE). DOE guidebooks (e.g. DOE, 1993; DOE, 2009) define methods for preventing and mitigating human error for use by QA practitioners, including concurrent dual verification: "A series of actions by 2 individuals, at the same time and place, to separately confirm the condition before, during, and after an action, when the consequences of an incorrect condition would subsequently lead to undesired harm." The motivation for this study is to intersect the complementary disciplines of human factors and quality assurance, with a focus on the concurrent dual verification context of high consequence environments. The literature provides no evidence of a JPA approach in a QA context having ever been studied before, therefore this study fills a research void and should draw interest from a variety of disciplines.

METHODOLOGY

The primary emphasis of the original research question is on the effectiveness of a JPA, however there must be a task for the QA observers to witness such that they can perform the role of concurrent dual verification. Basic guidelines for selecting this task were as follows:

- Not be too complex or time-consuming
- Not be too simple, such that the ability to inject faults would be difficult as they would be too obvious
- Be consistent with a task that may be conducted in a high consequence manufacturing environment

A predefined Lego™ assembly task was determined to satisfy these conditions, with a corresponding checklist for the QA observer to follow. An advantage of selecting a Lego™ assembly task over more applied techniques (such as repairing a lawnmower engine) is that there is not any built-in covariance of prior knowledge for participants; in other words, they would all be novices. Another advantage is the similarity of the Lego™ task with those conducted in a high consequence manufacturing environment. Assemblers are often provided with a kit of similar-looking parts (e.g., fasteners of different size but equal length) and instructions how to install them, and in many real-world applications there is a QA observer to oversee the task and ensure that no errors are made. A within-subjects design model was chosen so this drove the need for a second assembly task. It was not desirable to perform the same task again due to the potential for learning effects, therefore it was determined that two different Lego™ assembly tasks (Pattern A, Pattern B) were appropriate and faults could be introduced by the inclusion of incorrect pieces into the patterns. Another variable that should be reasonably controlled is the difference in size or complexity between the two patterns. This constraint eliminated many of the popular Lego™ patterns from contention, such as Star Wars™ or super heroes, since the ability to equalize both size and complexity across two patterns is difficult. Further, defined patterns with uniquely shaped pieces require all parts to be used during assembly (i.e., there are no pieces left over). This might make it too obvious to inject faults into these complex patterns.

Pattern A has 104 pieces and Pattern B has 150 pieces, with 7 pieces in each pattern specifically chosen for inserting a fault. Since one of the constraints is to avoid selecting a task that is too simple, the patterns were not assembled as stand-alone kits but instead contained within a larger set of pieces that may or may not be used in either sub-assembly. This eliminates the potential confound where the checker might notice that an "incorrect" part was used if there are supposed to be no pieces remaining when the assembly task is completed. The total number of available pieces needs to be higher than both sub-assemblies combined, with additional margin such that there is sufficient uncertainty in the experiment to warrant the use of a JPA. Since the total number of pieces in Pattern A and Pattern B is 254 (almost $256 = 2^7$ or 7 bits of information), a total assembly kit of 512 pieces was chosen since $512 = 2^9$ or 9 bits of potential information (Posner, 1964). However, the total number of bits of information is much higher than 9 bits when considering the different part shapes (23 for Pattern A), colors (6 for Pattern A), and markings that each act as multipliers to the total number of possible combinations of parts within the two patterns. This large number is roughly doubled when adding in the effects of Pattern B, and doubled again when considering parts not used in either pattern. Thus there is tremendous (and sufficient) uncertainty in the kit of 512 parts that experimental participants were not able to determine the total number of pieces being used for each pattern. If the JPA is designed to improve the chances that the *correct piece* in the *correct color* and *correct shape* is installed in the *correct way*, at the *correct time* and *correct place* in the assembly process, then it would be a tremendously useful verification tool for mitigating human error.

Faults were inserted by the assembler into both patterns in specific pre-determined locations, with the pre-test instruction that the participant observe the assembly task and tell the assembler if they noticed any errors. The fault types are:

1. Insert the incorrect piece that has markings. For example, instead of inserting a 2 x 2 yellow piece another 2 x 2 yellow piece with a pattern on one side is installed.
2. Insert the incorrect piece(s) but with no markings. For example, instead of inserting a 2 x 4 black piece two 1 x 4 black pieces are installed.
3. Insert the correct piece(s) but in the incorrect configuration. For example, instead of constructing a 2 x 8 piece on the left and a 2 x 2 piece on the right, the order is switched and the 2 x 2 piece goes on the left
4. Insert the correct piece(s) but in the wrong location or orientation. For example, a window piece is installed backwards.

The fault types were chosen for their reasonable similarity to those that occur in high consequence environments. For example, fault type 4 is similar to a component being installed backwards on a printed circuit board. The specific faults in the assembly task were only selected for their ease of insertion and ability to avoid detection by the QA checker, and were not equally spaced throughout the assembly task. For one of the two assembly tasks in each experimental trial, the JPA was provided for review beforehand and was available throughout the duration of that assembly task. Table 1 summarizes how the order of observation varied across the 24 participants. By varying the order of assembly in this way, specific effects of the presence of a JPA (if they exist) can be distinguished from the effects of sequence of assembly.

Table 1: Grouping of Experimental Participants By Sequence of Assembly

Number of Participants	Sequence of Assembly, Presence of JPA	Abbreviation
6	Pattern A without JPA, followed by Pattern B with JPA	A{JB}
6	Pattern A with JPA, followed by Pattern B without JPA	{JA}B
6	Pattern B without JPA, followed by Pattern A with JPA	B{JA}
6	Pattern B with JPA, followed by Pattern A without JPA	{JB}A

The number of participants was determined from the estimated probability of detecting each fault, with different values estimated for with and without a JPA. The simplifying assumption was to assign a constant average probability across all fault instances. Since the checklist specifically identified the 4 different fault types, it was assumed that a reasonable average probability of detection was p_1=0.5 without a JPA and p_2=0.9 with a JPA. The number of experimental participants (n) must be a multiple of 4 to match the test conditions in Table 1 and therefore make the experiment balanced. For n*14 binary trials (n*7 with a JPA and n*7 without a JPA), the probability of concluding that there is a difference when p_1=p_2 depends on n as follows in Table 2:

Table 2: Probability of Correctly Concluding that p2>p1

n	Type-1 error of 0.05	Type-1 error of 0.025
4	0.96	0.93
8	0.9995	0.998
12	~1	0.99998
≥16		~1

Because of simplifying assumptions, 24 participants were chosen to ensure adequate margin. Candidates who perform QA activities as part of their normal job were specifically excluded from this study, so all checkers were novice users in this role. The JPA for this experiment (see Figure 1) consists of a short, concise, and simple checklist intended to elicit behavioral cues that would enhance the detection of faults in this experiment.

- Your role as an observer is an essential part of this important task. Complex assemblies require a second set of eyes in order to catch any errors.
- Pay attention for the following types of error:
  - An incorrect piece is installed, meaning that it is either the wrong size, wrong color, or wrong markings
  - The correct piece is installed, but in the wrong orientation
  - The correct piece is installed, but in the wrong location
- Feel free to ask questions about the task at any time. If necessary, ask the assembler to stop until you are comfortable with proceeding.
- The assembler should not turn to the next page of the instructions without your approval.
- For each page of the instructions, the order of assembly does not matter.
- The box contains 512 total parts. Some parts will be used and some will not.

Figure 1: Job Performance Aid

Each of the six checklist items was considered to be essential information, with the following rationale:

- *Your role as an observer is an essential part of this important task.* It is the author's experience that novice QA checkers do not always recognize the importance of simply being an observer for an important task, and subject matter experts who perform the work may sometimes resent a non-expert "checking their work."
- *Pay attention for the following types of error.* Clark (1982) and White et al. (2010) recommend that high-risk concerns be specifically identified, and to avoid confusion the 4 fault types were addressed in only 3 statements.
- *Feel free to ask questions about the task at any time.* This statement is to build the QA checker's confidence by signaling that they are vital to the assembly process and could stop it at any time.
- *The assembler should not turn to the next page of the instructions without your approval.* This statement was necessary to categorize participant response data.
- *For each page of the instructions, the order of assembly does not matter.* This statement was necessary to reduce false alarms in the data.
- *The box contains 512 total parts. Some parts will be used and some will not.* This statement eliminates the confound whereby the QA observer could use the presence or absence of parts on the table as a cue.

Clark (1982), Shriver et al. (1982), and Smillie (1985) are emphatic that the key final step in successful development of a JPA is verification and validation with expert users, therefore a pilot study was performed on this JPA with 4 SNL QA experts. The process relies heavily on these reviews to identify and correct procedural ambiguities, omissions, and inaccuracies.

RESULTS AND ANALYSIS

14 binary observations were recorded for each participant and are summarized in Table 3. Participant #7 was disqualified due to misunderstanding the pre-test instructions and therefore replaced with participant #25. The three Pattern B instances where only 6 trials appeared were from wrongdoing by the assembler and a fault could not be inserted. Note that participant #1 detected every fault in both patterns.

Table 3: Fault Detection Performance, By Participant

Subject	Pattern A Trials	Pattern A Detections	Pattern B Trials	Pattern B Detections	Percent Detected
1	7	7	7	7	100%
2	7	5	6	5	77%
3	7	4	7	3	50%
4	7	7	6	3	77%
5	7	6	6	4	77%
6	7	5	7	4	64%
8	7	4	7	5	64%
9	7	6	7	5	79%
10	7	5	7	7	86%
11	7	3	7	4	50%
12	7	3	7	4	50%
13	7	3	7	4	50%
14	7	3	7	6	64%
15	7	4	7	4	57%
16	7	4	7	5	64%
17	7	6	7	3	64%
18	7	3	7	4	50%
19	7	4	7	2	43%
20	7	4	7	3	50%
21	7	4	7	5	64%
22	7	4	7	5	64%
23	7	4	7	5	64%
24	7	3	7	3	43%
25	7	4	7	3	50%

Table 4 summarizes the results for each particular fault. The 3 instances of wrongdoing by the assembler are seen again where two of the trials associated with fault #8 and one trial with fault #11 were deemed to be a "no test". It is noteworthy that fault #1, fault #6, fault #7, fault #10, and fault #14 were frequently missed and are all marking faults (fault type 1). Fault #12 (also fault type 1) was specifically designed to be noticed and, as expected, frequently detected.

Table 4: Performance, By Fault Number

Pattern	Fault Number	Fault Type	Number of Trials	Number of Detects	Percent Detected
A	1	1	24	5	21%
A	2	3	24	24	100%
A	3	3	24	23	96%
A	4	3	24	24	100%
A	5	4	24	17	71%
A	6	1	24	6	25%
A	7	1	24	6	25%
B	8	2	22	15	68%
B	9	4	24	21	88%
B	10	1	24	5	21%
B	11	3	23	23	100%
B	12	1	24	20	83%
B	13	2	24	17	71%
B	14	1	24	2	8%

Binary logistic regression (Agresti, 2013) was used to model the probability of detecting a fault as a function of the experimental factors. For Pattern A, faults #2 and #4 were excluded from the models because they were always detected; for Pattern B, fault #11 was excluded for the same reason. The model form used in each case is:

$$log\left(\frac{\pi(Err(i),\ Seq(j))}{1-\pi(Err(i),\ Seq(j))}\right) = \alpha_0 + \beta_i + \gamma_j, \qquad (1)$$

where α_0 represents the log odds at a standard experimental condition, β_i reflects the change in log odds when changing the experimental condition from the standard fault number to fault #*i*, and γ_j reflects the change in log odds when changing the experimental condition from the standard sequence to sequence *j*. In the case of both patterns, the standard sequence is denoted by A{JB} (see Table 1).

Table 5 and Table 6 display the parameter estimates related with patterns A and B, respectively.

Table 5: Logistic Regression Table, Pattern A

Parameter	Estimate	SE Estimate	Z-ratio	P-value
α_0	-2.845	0.810	-3.51	0.000
$\gamma_{B\{JA\}}$	1.792	0.776	2.31	0.021
$\gamma_{\{JA\}B}$	1.999	0.778	2.57	0.010
$\gamma_{\{JB\}A}$	1.578	0.775	2.04	0.042
β_3	4.967	1.218	4.08	0.000
β_5	2.494	0.731	3.41	0.001
β_6	0.251	0.710	0.35	0.724
β_7	0.251	0.710	0.35	0.724

Table 6: Logistic Regression Table, Pattern B

Parameter	Estimate	SE Estimate	Z-ratio	P-value
α_0	0.893	0.593	1.50	0.132
$\gamma_{B\{JA\}}$	0.350	0.637	0.55	0.582
$\gamma_{\{JA\}B}$	-0.615	0.610	-1.01	0.313
$\gamma_{\{JB\}A}$	-0.187	0.612	-0.31	0.760
β_9	1.211	0.776	1.56	0.119
β_{10}	-2.150	0.691	-3.11	0.002
β_{12}	0.870	0.722	1.20	0.228
β_{13}	0.132	0.650	0.20	0.839
β_{14}	-3.227	0.879	-3.67	0.000

With a Hosmer-Lemeshow value of p=0.725, there is no evidence for lack of fit in equation (1) and thus the model is reasonably accurate.

DISCUSSION

The effect of the JPA can be deduced by considering the complete set of estimated γ_j terms in the model, which represent the 4 test conditions shown in Table 1. In the case of pattern A (see Table 5), note that $\hat{\gamma}_{B\{JA\}}$, $\hat{\gamma}_{\{JA\}B}$, and $\hat{\gamma}_{\{JB\}A}$ are all statistically significantly non-zero and positive and thus imply increased probability of detection versus the standard sequence A{JB}. This means that faults in Pattern A were detected *less frequently* in the standard A{JB} sequence, suggesting that the JPA may have had only a limited effect in this experiment. The only distinguishable effect of the JPA is its presence in a 3-way interaction between sequence, presence/absence of a JPA, and Pattern A. If there was an effect from *only* the JPA, the $\hat{\gamma}_{B\{JA\}}$ and $\hat{\gamma}_{\{JA\}B}$, terms in Table 5 would be statistically significant and positive while $\hat{\gamma}_{\{JB\}A}$ would be near zero. In contrast, the effects of sequence and/or presence/absence of a JPA on the probability of detection for pattern B (see Table 6) were not observed.

As seen in Table 4, marking faults (fault type 1) were frequently missed and therefore dominate the 3-way interaction term in the results. This can be seen in the β_3 term in Table 5 ($p < 0.0005$) with a large positive value for the estimate (4.967), indicating that for Pattern A participants were much more likely to detect fault #3 (also fault type 3) compared to fault #1 (a marking fault). This suggests that the reason Pattern A appears in the 3-way interaction term is because it has more marking errors and thus gives a better opportunity to detect differences in the probability of detection between the different fault types.

CONCLUSION AND FUTURE RESEARCH

The research presented in this thesis is the first known example of evidence-based job performance aid (JPA) use in a quality assurance setting. This study used an assembly task and checklist format as the vehicle for the experiment, and incorporated previously recognized design principles from the literature into the configuration of the JPA. The design of the task included some simplifying assumptions regarding an average probability of detection between different types of error, and this led to the primary research hypotheses unable to be proved. Nevertheless, there were 3 important findings.

First, this study created a testing methodology sensitive enough to detect differences in the effects on performance between the 1) sequence of observation of patterns, 2) presence/absence of a JPA, and 3) Pattern A. The author reflects that if the *main effect* of a JPA on performance of a concurrent dual verification task were easily identifiable then it would likely have been detected long ago. Second, the results indicate that concurrent dual verification *itself* is not necessarily an effective control. One cannot assume that having a QA checker in place will have an impact on error detection or mitigation, especially for specific types of errors. Third, the assumption of average probability of detection between different types of error may have been demonstrably wrong but the error detection probability could be empirically *verified* within the error construct created during this experiment. The recognition in this study that different types of error have different probabilities of detection, which could then be inserted into a checklist as behavioral cues, might be highly useful as a leverage to the success of both JPA design and concurrent dual verification activities.

Both the methodology and results of this study are an effective baseline from which to launch future research activities. Repeating the same experiment with a uniform fault type, which would provide a constant probability of detection for all faults, may yield intriguing results. A focus on marking faults, which can be considered signals, might provide an attractive opportunity. Since incorrect markings may fall below the signal detection threshold (Swets, 1964), such a focused study might extend this essential paradigm of vigilance and attract a wide audience. If a similar experiment equally spaced the faults (possibly by elapsed time) throughout the assembly task, then perhaps a vigilance decrement could be studied in a number of ways. These and other vigilance studies may benefit from the recognition of different probabilities of detection for different types of error, both within and outside of the concurrent dual verification construct. Finally, it is worth noting that the QA checkers in this experiment used a simple checklist and thus future studies may have different results with other JPA formats. Perhaps as a result of such studies, optimal JPA formats for different error types may emerge.

REFERENCES

Agresti, A. (2013). *Categorical Data Analysis* (3rd ed.). New York: John Wiley.

Clark, J.E. (1982). *JPAs: An Actual Experience*. UNI-SA-96; CONF-820223-1. Richland, WA: UNC Nuclear Industries, Inc.

DOE. (1993). *Guide to Good Practices for Independent Verification*. DOE-STD-1036-93. Washington, DC: United States Department of Energy.

DOE. (2009). *Human Performance Improvement Handbook (Vol. 2).* DOE-HDBK-1028-2009. Washington, DC: Department of Energy.

Folley, J.D., Jr. (1961). *A Preliminary Procedure for Systematically Designing Job Performance Aids*. Technical Report 61-550. Wright-Patterson AFB, OH: Air Force Systems Command.

Folley, J.D., Jr., Joyce, R.P., Mallory, W.L., & Thomas, D.L. (1971). *Fully Proceduralized Job Performance Aids: Volume 1. Draft Specification for Organizational Maintenance.* Technical Report AFHRL-TR-71-53. Wright-Patterson AFB, OH: Air Force Human Resources Laboratory.

Gawande, A. (2010). *The Checklist Manifesto: How to Get Things Right*. New York: Picador.

Goff, J., Schlesinger, R., & Parlog, J. (1969). *Project PIMO Final Report. Volume 2. PIMO Test Summary*. Chatsworth, CA: Serendipity, Inc.

Haynes, A.B., … & Gawande, A.A. (2009). A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population. *New England Journal of Medicine*, *360*(5), 491-499.

Helmreich, R.L., & Merritt, A.C. (2001). *Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences*. Farnham, UK: Ashgate Publishing.

Joyce, R.P., Chenzoff, A.P., Mulligan, J.F., & Mallory, W.J. (1973a). *Fully Proceduralized Job Performance Aids: Vol. I. Draft Military Specifications for Organizational Level and Intermediate Maintenance*. Technical Report AFHRL-TR-73–43(I). Wright-Patterson AFB, OH: Air Force Human Resources Laboratory.

Joyce, R.P., Chenzoff, A.P., Mulligan, J.F., & Mallory, W.J. (1973b). *Fully Proceduralized Job Performance Aids: Vol. II. Handbook for JPA Developers*. Technical Report AFHRL-TR-73–43(II). Wright-Patterson AFB, OH: Air Force Human Resources Laboratory.

Mackie, R.R. (1964). *Six Years of Research on Human Factor Problems in ASW.* Technical Report 206-28. Los Angeles, CA: Human Factors Research Inc.

Mackworth, N.H (1950). Researches on the Measurement of Human Performance. *Medical Research Council Special Report Series,* No. 268.

Mattson, R.J. *An Assessment of Criticality Safety at the Department of Energy Rocky Flats Plant*. SCIE-DOE-201-89, July-September, 1989.

McKenzie, R.M. (1958). On the Accuracy of Inspectors. *Ergonomics 1*(3), 258-272.

Meilinger, P.S. (2004). When the Fortress Went Down. *Air Force Magazine, 87*, 78-82.

Miller, R.B. (1953). *A Method for Man-Machine Task Analysis*. Report 53-137. Wright-Patterson AFB, OH: Wright Air Development Center.

Miller, R.B. (1956). *A Suggested Guide to the Preparation of Handbooks of Job Instructions*. Pittsburgh, PA: American Institute for Research.

Newman, S.E. (1957). *On Identifying Effective Handbook Techniques* (ML-TM-57-26). Lowry Air Force Base, CO: Air Force Personnel and Training Research Center.

Norman, D.A. (1981). Categorization of Action Slips. *Psychological Review, 88*(1), 115.

Posner, M.I. (1964). Information Reduction in the Analysis of Sequential Tasks. *Psychological Review, 71*(6), 491-504.

Rasmussen, J. (1982). Human Errors: A Taxonomy for Describing Human Malfunction in Industrial Installations. *Journal of Occupational Accidents, 4*(2), 311-333.

Reason, J. (1990). *Human Error*. London, UK: Cambridge University Press.

Russians Miswire Satellite. (1996, March 21) *The Moscow Times*. Retrieved from http://www.themoscowtimes.com/news/article/russians-miswire-satellite/326829.html

See, J.E. (2012) *Visual Inspection: A Review of the Literature*. No. SAND2012-8590. Sandia National Laboratories.

Shriver, E.L. (1975). *Fully Proceduralized Job Performance Aids: Guidance for Performing Behavioral Analyses of Tasks*. Falls Church, VA: URS/Matrix Research Company.

Shriver, E.L., Zach, S.E., and Foley Jr, J.P. (1982). *Test of Job Performance Aids for Power Plants*. No. EPRI-NP-2676. Alexandria, VA: Kinton, Inc.

Smillie, R.J. (1985) Design Strategies for Job Performance Aids. In T.M. Duffy & R.W. Waller (Eds.) *Designing Usable Texts*, 213-241. Academic Press, Inc.

Swets, J.A. (1964). *Signal Detection and Recognition by Human Observers*, New York: Wiley & Sons.

Tsao, Y.C., Drury, C.G., & Morawski, T.B. (1979). Human Performance in Sampling Inspection. *Human Factors, 21*(1), 99-105.

White, R.E., Trbovich, P.L., Easty, A.C., Savage, P., Trip, K., & Hyland, S. (2010). Checking it Twice. *Quality and Safety in Health Care, 19*(6), 562-567.

Wiener, E.L. (1975). Individual and Group Differences in Inspection. In C.G. Drury & J.G. Fox (Eds.) *Human Reliability in Quality Control*, 101-122. Halsted Press.