SAND2016-1424C

Exceptional service in the national interest

Sandia National Laboratories



The Effect of Job Performance Aids on Quality Assurance

Erik Fosshage, Sandia National Laboratories

NASA Quality Leadership Forum

March 9-10, 2016

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2011-XXXXP

Motivation

- Quality Engineer for Sandia National Laboratories since 2005
- Purdue MSIE, May 2014, Human Factors Engineering
- Wanted to bridge the disciplines of Human Factors and Quality Assurance (QA)
- Previously created a ***job performance aid*** (JPA) for novice QA co-workers for concurrent dual verification tasks
- A checklist is one type of JPA (others are procedures, manuals, training videos, etc.)
- First-ever research on JPAs in a QA context

Quality Assurance Context

- DOE *Guide to Good Practices for Independent Verification (1993):*

  Concurrent Dual Verification – A method of checking an operation, an act of positioning, or a calculation in which the verifier independently observes and/or confirms the activity

- NASA-STD 8709.22 (2010) definitions:

  Process Witnessing – Physical observation of a contractor test or work process to ensure that the process is being correctly performed in accordance with prescribed procedures and contract requirements.

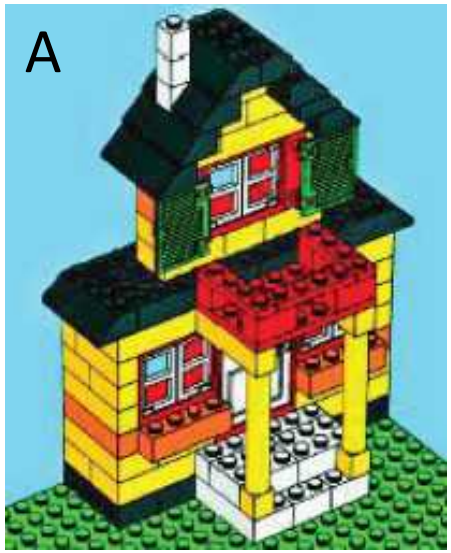
History

- Boeing 299 (later B-17) crash in 1935 led to pilot's checklist
- USAF behavioral research on training aids (e.g. Miller, 1953) led to the "Task Analysis" methodology
- JPA research continued through the 1970s; findings included:
  - Reduced errors in complex tasks that were infrequently performed
  - Shortened the training time for novice users
  - Different formats (pictures or text) conveyed information differently
- JPA interest resurfaced after Three Mile Island incident (1979)
- JPAs now adopted by various "high consequence" industries: aviation, nuclear power, medicine, aerospace
- Popular interest: *The Checklist Manifesto* (2010)

Experimental Task Selection

- Guidelines:
  - Not too simple, not too complex
  - Consistent with high consequence environment
- Solution: Lego™ assembly task
  - Participant expertise not a covariant: all users are novices
  - Reasonable similarity to manufacturing environment
  - Easy to inject faults and measure performance
- Within subjects design, 2 different Lego™ patterns
  - One assembled with JPA present, one assembled without
  - 24 participants, counterbalanced for learning effect

Lego™ Patterns



Pattern A: 104 pieces

Pattern B: 150 pieces

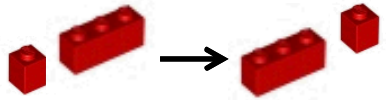
7 faults injected into each pattern (14 total)

Fault Types:

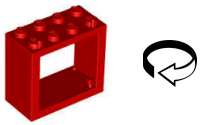
1. Markings

2. Incorrect piece(s)

3. Wrong order



4. Wrong orientation



Assumption: Constant probability of detection for all fault types

JPA Design

- Common themes in the literature*:
  - The focus should be on the user
    - Fully understand the job function
    - Fully understand the behaviors used
  - Information must be task oriented
    - Brief, concise, explicit instructions; be directive and action-specific
    - Use simplified and standard language
  - Final important step: validation with expert users
- JPA for this experiment:
  - Short, concise, and simple checklist
  - Elicits behavioral cues to enhance the detection of faults

*** Best references are Shriver et al. (1982), Smillie (1985), and Gawande (2010)**

Checklist

- 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.

Behavior cues
Error avoidance

Results (1)

- Participant scores ranged from 43% - 100% detection of faults
  - Majority of participants scored in the 50-60% range
  - Traditional inspection results yield ~80% success rate
- Poor performance overall
- Suggests limitations to concurrent dual verification

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%
5	7	7	6	3	77%
6	7	6	6	4	77%
7	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%

Results (2)

- Performance by fault number (and fault type) yielded more intriguing results
  - Faults 2, 4, and 11 were always detected (type 3, wrong order)

  - Fault type 1 (markings) frequently missed

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%

Marking errors (fault type 1) are more difficult to detect

Analysis (1)

- Binary logistic regression (Agresti, 2013) used to model the probability of detecting a fault

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

- Estimates for Pattern A
  - γ terms are all statistically non-zero and *positive*
  - Faults detected *less frequently* in the standard sequence:
    - A{JB}, or…
    - Pattern A first, then Pattern B with checklist

Parameter	Estimate	Standard Error 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

3-way interaction between sequence, checklist presence, and Pattern A

Analysis (2)

- Estimates for Pattern A
  - Fault #3 (incorrect order) detected *more frequently* than the standard fault #1 (markings)
  - Same effect for β_5, which is a wrong orientation fault

Parameter	Estimate	Standard Error 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

This suggests that Pattern A appears in the 3-way interaction because it has more marking errors

Fitted Model Validation

- No evidence for lack-of-fit in the model
- Formal tests (where p > 0.05 is significant):
  - Pearson: p=0.171
  - Deviance: p=0.194
  - Hosmer-Lemeshow: p=0.725
- Reasonable similarity between Estimated Probability of Detection and Observed Fraction of Detection
- However…

Fault #	Sequence	Estimated Probability of Detection	Observed Fraction Detected
1	A {JB}	0.055	0.000
1	B {JA}	0.259	0.500
1	{JA} B	0.300	0.167
1	{JB} A	0.220	0.167
3	A {JB}	0.893	1.000
3	B {JA}	0.980	1.000
3	{JA} B	0.984	1.000
3	{JB} A	0.976	0.833
5	A {JB}	0.413	0.500
5	B {JA}	0.809	0.500
5	{JA} B	0.839	0.833
5	{JB} A	0.773	1.000
6	A {JB}	0.069	0.000
6	B {JA}	0.310	0.333
6	{JA} B	0.355	0.500
6	{JB} A	0.266	0.167
7	A {JB}	0.069	0.000
7	B {JA}	0.310	0.333
7	{JA} B	0.355	0.333
7	{JB} A	0.266	0.333

The probability of detection for each fault is **not** equal.

Finding (1)

- Created a testing methodology sensitive enough to detect differences in the effects on performance between:
  - Pattern sequence
  - Checklist presence
  - Pattern A
- If the *main effect* of a checklist on performance (of a concurrent dual verification task) were easily identifiable, then it would have been detected long ago

Finding (2)

- The assumption of average probability of detection between different types of error was ***empirically verified*** to be wrong

- Fault (Error) Types:
  1. Markings
  2. Incorrect piece(s)
  3. Wrong order
  4. Wrong orientation

Finding (3)

- Concurrent dual verification is not necessarily an effective control against defects, both ***with*** and ***without*** a checklist

- Verification techniques presented in the literature may be *conditional,* especially for specific types of errors (ie: markings)
- No JPA format is best for all circumstances
- Quality assurance tools must be well designed and well understood by ***both*** the designer and the user, in order to effectively control risk

Conclusions

Sandia National Laboratories

- This is the first known research study to have examined:
  - The effect of a checklist on performance in a quality assurance setting
  - Subtle and complex interactions between JPA design, error types, and base error probability of detection
  - Probability of detection of different error types in the following context:
    - Quality Assurance (concurrent dual verification)
    - Use of a JPA, specifically a checklist
    - Simple assembly task

References

- Agresti, A. (2013). *Categorical Data Analysis* (3rd ed.). New York: John Wiley.
- DOE. (1993). *Guide to Good Practices for Independent Verification.* DOE-STD-1036-93. Washington, DC: United States Department of Energy Technical Standards Program.
- Fosshage, E. (2014). *The Effect of Job Performance Aids on Quality Assurance* (No. SAND2014-4762). Albuquerque, NM: Sandia National Laboratories.
- Gawande, A. (2010). *The Checklist Manifesto: How to Get Things Right*. New York: Picador.
- Kemeny, J.G. (1979). *The Need for Change, the Legacy of TMI: Report of the President's Commission on the Accident at Three Mile Island*. Washington, DC: U.S. Government Printing Office.
- Miller, R.B. (1953). *A Method for Man-Machine Task Analysis*. Technical Report 53-137. Wright-Patterson AFB, OH: Wright Air Development Center.
- NASA. (2010). *Safety and Mission Assurance Acronyms, Abbreviations, and Definitions.* NASA-STD 8709.22. Washington, DC: National Aeronautics and Space Administration.
- 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.