SAND2014-2324P

Exceptional service in the national interest

Sandia National Laboratories

PURDUE ENGINEERING

The Effect of Job Performance Aids on Quality Assurance

Erik Fosshage

Master's Thesis Defense

Purdue University – March 24, 2014

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

- Looked for a way to bridge the disciplines of Human Factors and Quality Assurance
- Previous incident in the author's past led to the creation of a job performance aid (JPA) for a novice QA co-worker in a concurrent dual verification context
- JPA research builds on work from Stevens-Adams et al. (2013) regarding the effectiveness of verification techniques
- Dr. Caldwell immediately recognized that JPAs for QA had probably not been previously studied
- Shriver et al. (1982) intended to design JPAs for 8 job functions in a power plant, however the JPA for QA was later dropped from the study.

Literature Review

- 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 "Task Analysis" methodology
  - Instructions are most effective when incorporating behavioral "cues"
  - Clear, concise, user-focused statements
- JPA research continued through the 1970s
  - Reduce errors in complex tasks that were infrequently performed
  - Shorten 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* (Gawande, 2010)

SNL Quality Assurance Context

- Sandia National Laboratories
  - Albuquerque, NM and Livermore, CA
  - Specializes in the design and production of high consequence products for US government customers, primarily nuclear weapons
  - Oversight provided by the Department of Energy (DOE)
- DOE *Human Performance Handbook* (2009) categories:
  - Self checking
  - Peer checking
  - Peer review
  - Concurrent verification
  - Independent verification
- 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

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
    - Kit of fasteners with different sizes but equal length; orientation markings
    - QA observer often present in high consequence assemblies
  - Easy to inject faults and measure performance
  - Instructions already printed
- Within subjects design, 2 different Lego™ patterns
  - Pattern A (104 pieces), Pattern B (150 pieces)
  - One assembled with JPA present, one assembled without
  - 7 faults injected into each pattern (14 total)

Lego™ Patterns

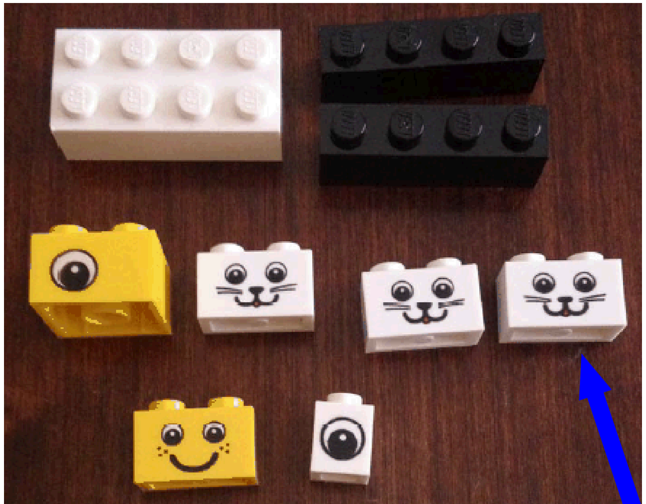

Pattern A

- 104 pieces
- 23 shapes
- 6 colors


Pattern B

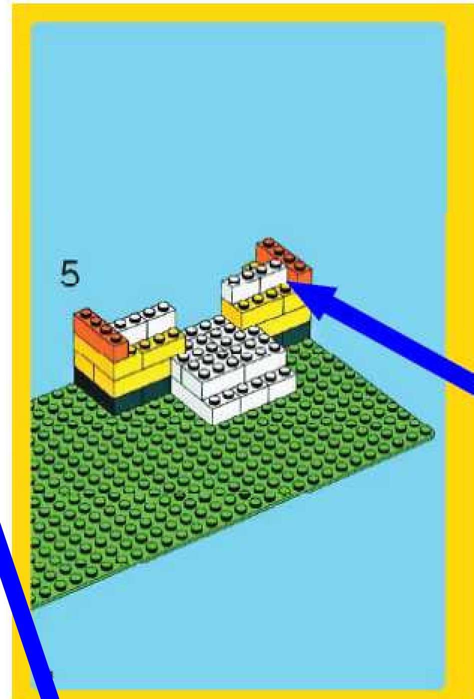
- 150 pieces
- 32 shapes
- 9 colors

Lego™ Faults



Fault Types:

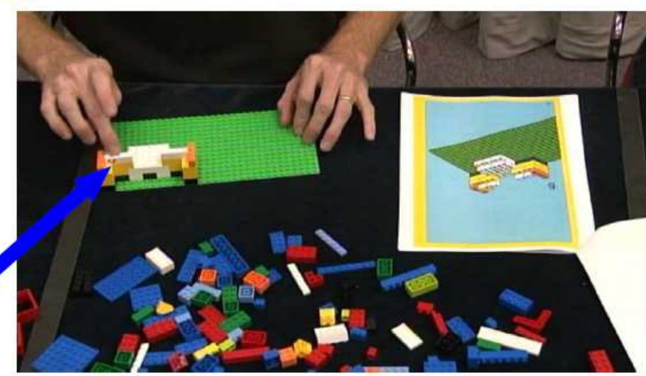
1. Markings
2. Incorrect piece(s)
3. Wrong order
4. Wrong orientation



Fault Type 1 Example

white 1 x 2 piece

Replaced by white 1 x 2 piece with markings



Definitions

Fault	An intentional and specific error that is introduced by the assembler
Mistake	An unintentional error by the assembler
Error	The use of an incorrect piece or incorrect assembly order
Miss	An error that is not noticed by the checker; includes catches
Detection	Before turning the next page of the assembly instructions, the checker identifies an error
Catch	An error that is noticed by the checker later (after the page has been turned)
False alarm	Any response from the checker when there is no error present
Violation	A fault that is purposely hidden from the checker, such that there is not an opportunity to notice it
Sequence error	The checker verbally authorizes the assembler to turn the page of the instructions before the assembly task on that page has been completed; treated as a false alarm

- Sequence error considered a mode error (Norman, 1981)

Experimental Variables

- Independent variables:
  - Pattern sequence (A first or B first)
  - JPA (present or not present)
- Uncontrolled variables (potential covariants):
  - Pattern A elapsed time
  - Pattern B elapsed time
  - Total elapsed time
  - Fault type (1, 2, 3, or 4)
- Dependent variables
  - Detections
  - Misses (includes catches)
  - False alarms (includes sequence errors)

Primary Experimental Hypothesis

H_1: The presence of a JPA has no effect on the detection of faults in the quality assurance role.

H_{1A}: The presence of a JPA ***has*** an effect on the detection of errors in the quality assurance role.

Other Experimental Hypotheses

- Addresses Independent variables:
  - Pattern sequence (A first or B first)
  - JPA (present or not present)

H_2: There is no difference in the detection of faults with a JPA than without a JPA.

H_3: The order of presentation of the JPA has no effect on the detection of faults in the quality assurance role.

H_4: The order of presentation of the different patterns has no effect on the detection of faults in the quality assurance role.

Test Approach

- Assumptions:
  1. Constant probability of detection for all fault types
  2. If JPA not present, probability of detecting fault: $p_1 = 0.5$
  3. If JPA present, probability of detecting fault: $p_2 = 0.9$
- For [n x 14]* binary trials, probability of concluding $p_2 > p_1$:

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 chosen
- Counterbalance for learning effect:

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

*** [n x 7] trials with a JPA + [n x 7] trials without a JPA**

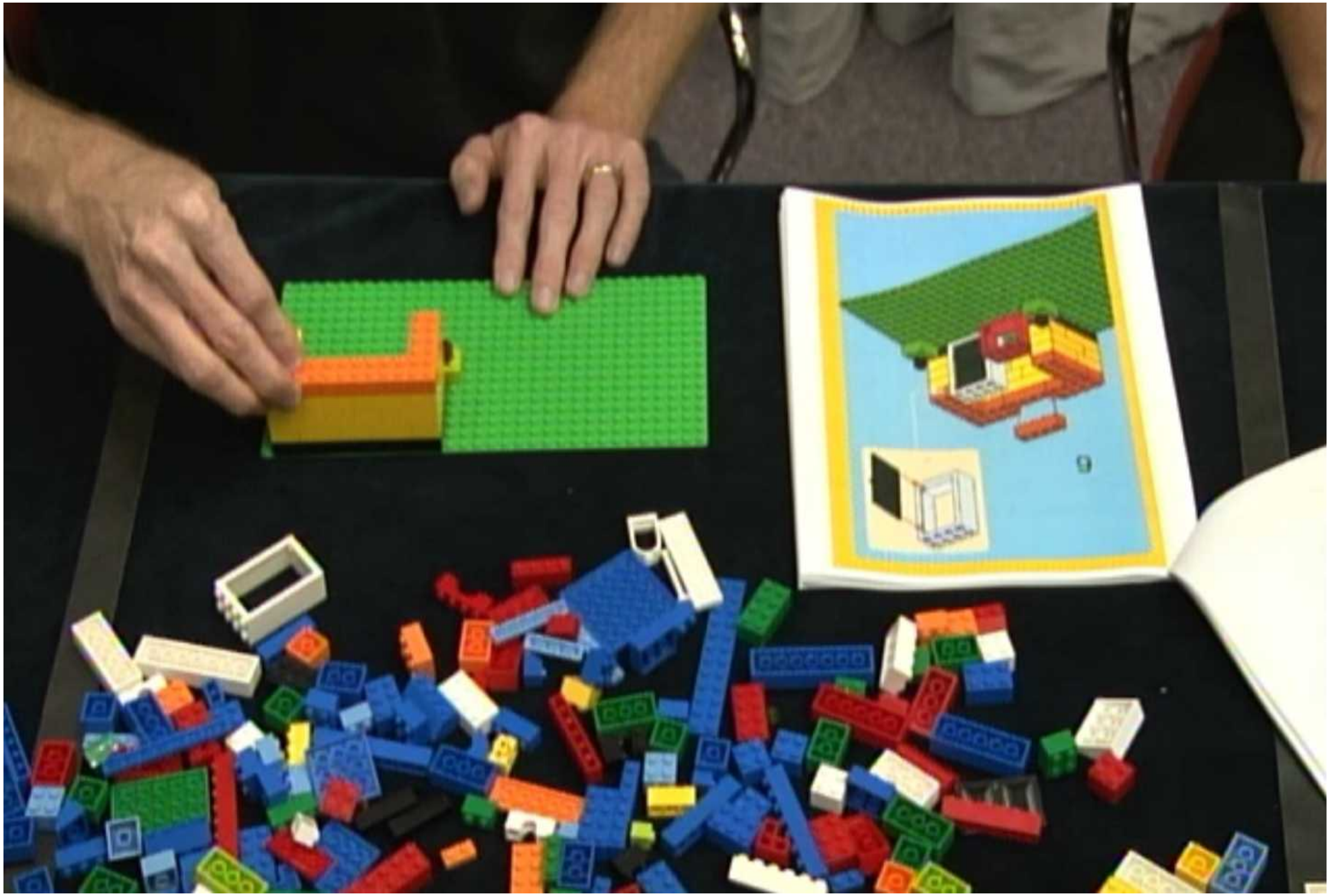
Additional Hypothesis

- Addresses assumption on the previous slide

H_5: The probability of detection for each fault is equal.

H_{5A}: The probability of detection for each fault is **not** equal.

Test Setup



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
    - Identify *exactly* what the user needs to do
    - Brief, concise, explicit instructions; be directive and action-specific
    - Use simplified and standard language
    - Include *only* information that is absolutely necessary
  - 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)**

Job Performance Aid

- 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.

Results

- Participant scores ranged from 43% - 100% detection of faults
- 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, except #12 that was designed to be detected
  - Assembler error resulted in 3 instances of “no test”

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

Statistical Model

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

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

- α_0: log odds at a standard experimental condition
  - Requires standard fault number (#1 for Pattern A, #8 for Pattern B)
  - Requires standard pattern sequence (Pattern A first for both patterns, or A{JB}
- β_i: change in log odds when changing the standard fault number
- γ_j: change in log odds when changing the standard sequence

Analysis (1)

- Estimates for Pattern A
  - γ terms are all statistically non-zero and *positive*
  - Faults detected *less frequently* in the standard sequence A{JB}
  - γ terms indistinguishable from *each other* (SE estimates equal)
  - Since $\gamma_{\{JB\}A}$ is non-zero, JPA has only a limited effect by itself

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

- Probability of detection of Pattern A faults is impacted by *both* independent variables (pattern sequence and JPA presence)

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

Analysis (2)

- Estimates for Pattern A
  - β_3 term is statistically non-zero and positive
  - Fault #3 (fault type 3, incorrect order) detected *more frequently* than the standard fault #1 (fault type 1, markings)
  - Same effect for β_5, which is fault type 4 (wrong orientation)

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

- Marking errors (fault type 1) are more difficult to detect
- This suggests that the reason Pattern A appears in the 3-way interaction 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

Fitted Model Validation

- Model validation occurs by inverting equation (1) and replacing the parameters with their estimates. For fault #1 using the standard sequence A{JB}:

$$\hat{\pi}(Err(1), Seq(\text{A\{JB\}})) = \frac{exp(\hat{\alpha}_0)}{1 + exp(\hat{\alpha}_0)} = 0.055$$

…where $\hat{\alpha}_0 = -2.845$ (see previous slide)

- Similarly, for fault #6 and sequence B{JA}:

$$\hat{\pi}(Err(6), Seq(\text{B\{JA\}})) = \frac{exp\left(\hat{\alpha}_0 + \hat{\beta}_6 + \hat{\gamma}_{B\{JA\}}\right)}{1 + exp\left(\hat{\alpha}_0 + \hat{\beta}_6 + \hat{\gamma}_{B\{JA\}}\right)} = 0.310$$

- No evidence for lack-of-fit in the model where $p > 0.05$, therefore it is an accurate estimate
  - Pearson: p=0.171
  - Deviance: p=0.194
  - Hosmer-Lemeshow: p=0.725
- 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

H_{5A}: The probability of detection for each fault is **not** equal.

Experimental Hypotheses

- Not fully rejected, due to 3-way interaction between pattern sequence, JPA presence, and Pattern A:

  H1: The presence of a JPA has no effect on the detection of faults in the quality assurance role.

  H_2: There is no difference in the detection of faults with a JPA than without a JPA.

  H_3: The order of presentation of the JPA has no effect on the detection of faults in the quality assurance role.

  H_4: The order of presentation of the different patterns has no effect on the detection of faults in the quality assurance role.

Finding #1

- Created a testing methodology sensitive enough to detect differences in the effects on performance between:
  - Pattern sequence
  - JPA presence
  - Pattern A
- If the *main effect* of a JPA on performance (of a concurrent dual verification task) were easily identifiable, then it would have been detected long ago
  - Task analysis methodologies are limited by the difficulties in accurately predicting behavior due to the influence of cognitive factors that cannot be easily observed or modeled
  - A different experimental design may not have been able to detect the interaction between independent variables

Finding #2

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

Finding #3

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

- This is the first known research study to have examined:
  - The effect of a JPA 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, within the context of this experiment:
    - Quality Assurance (concurrent dual verification)
    - Use of a JPA, specifically a checklist
    - Simple assembly task

Future Research

- Redesigned study to counterbalance the learning effect
- Repeating the same experiment with a uniform fault type
  - Focus on marking faults only may yield intriguing results
  - Markings = signals --------> Signal Detection Theory
  - SDT, in turn, is an essential paradigm of vigilance theory
- More implications for vigilance theory
  - Equally spacing the faults throughout the experiment (number of pieces, or elapsed time) may allow study of vigilance decrement
  - Fault #12 (markings) was designed to be noticed, based on findings by Tsao, Drury, and Morawski (1979); later dropped from this study
  - Tsao and Wang (1984) later examined faults of different difficulty
- Future studies may have different results with other JPA formats

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.
- DOE. (2009). *Human Performance Improvement Handbook, Volume 2: Human Performance Tools for Individuals, Work Teams and Management.* DOE-HDBK-1028-2009. Washington, DC: United States Department of Energy Technical Standards Program.
- Gawande, A. (2010). *The Checklist Manifesto: How to Get Things Right*. New York: Picador.
- Miller, R.B. (1953). *A Method for Man-Machine Task Analysis*. Technical Report 53-137. Wright-Patterson AFB, OH: Wright Air Development Center.
- Norman, D.A. (1981). Categorization of Action Slips. *Psychological Review, 88*(1), 115.
- 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.
- Stevens-Adams, S., Cole, K., Wenner, C., & Dornburg, C. (2013). *The Effectiveness of Verification in Industrial Applications*. Unpublished manuscript, Sandia National Laboratories.
- Tsao, Y.C., Drury, C.G., & Morawski, T.B. (1979). Human Performance in Sampling Inspection. *Human Factors, 21*(1), 99-105.
- Tsao, Y.C., & Wang, T.G. (1984). Inspectors' Stopping Policy After Fault Detection. *Human Factors, 26*(6), 649-657.