NRC Research and/or Technical Assistance Rept

PDR

EGG-M-20382
PREPRINT

A Multimethods Approach to Safety Parameter Display Evaluation

W.W. Banks, H.S. Blackman, D.I. Gertman, R.J. Petersen

October 12-15, 1982

Tenth Water Reactor Safety Research
Information Meeting

EG&G Idaho
Idaho Falls, Idaho 83415

EGG-M-20382

A MULTIMETHODS APPROACH TO SAFETY PARAMETER DISPLAY EVALUATION

W. W. Banks, H. S. Blackman
D. I. Gertman, R. J. Petersen

EG&G Idaho, Inc.

The Human Factors Engineering Office of EG&G Idaho performed this NRC-funded study to assist the NRC in objectively assessing licensee-developed safety parameter display (SPD) formats and designs. The purpose of this study was to quantitatively measure the degree to which a tachistoscopic method of display evaluation would correlate with the results of a multidimensional rating approach to display evaluation. The ultimate goal was to identify the method which accounts for the greatest amount of operator performance, yet costs the least amount of money. Results of the following three experiments will be presented; (a) tachistoscopic, (b) multidimensional rating scale, and (c) the combined results of a and b.

The test material for all experiments consisted of three multivariate data display formats all under development as SPDs for reactor control rooms presenting safety parameter display data at the loss-of-fluid test (LOFT) facility. The three display formats studied were stars, deviation bar graphs, and meters. Three questions were posed: (a) What is the degree of concurrence between these two independent methods used in display evaluation? (b) Can one of the two methods be used successfully to predict results of the other? (c) What dimensions of SPD formats appear to be most crucial to operators for performance and preference?

Eighteen adult volunteers were used as subjects. Their ages ranged from 26 to 44 years and all reported vision correctable to 20/20. All were currently qualified reactor operators from the LOFT reactor plant, with a mean of 9.4 years reactor operating experience.

Tachistoscope Method. A dual-channel tachistoscope (t-scope) was used to study the three display formats. The classic model of signal detection was employed collecting data for perceptual sensitivity, response criterion, percent correct, and reaction time. Two studies were made: signal detection and parameter recognition. The signal detection study found differences for display type, exposure duration, and interactions. For the dependent variables of perceptual sensitivity, percent correct, and response criterion, stars were significantly greater than the combination of meters and bars, and stars were significantly greater than bars. The interactions of display and exposure duration also showed a superior performance for the star display, but only with the short exposure as the difference diminished with increasing exposure durations. Recognition study results revealed no significant effects or interactions from any of the analyses.

Multidimensional Rating Scale Method. The authors used a combination of factor-analytic and forced-choice techniques to develop six scales for evaluating display interfaces: content density, content integration, format, cognitive fidelity, cognitive processing, and general acceptance. The study sought to determine if this multidimensional rating scale (MDRS) methodology would apply to the evaluation of the three display formats.

Statistically significant results were obtained only for content integration (CI) and cognitive processing (CP). In both cases, the order of preference from most to least preferred was bars, star, and meters. Orthogonal planned comparisons showed that bars and star differed significantly from meters for CP only ($p < 0.05$); no other comparisons reached significance.

Combined Results. To answer the three major questions posed as objectives for this paper, forward stepwise multiple regression analyses were conducted. Two sets of analyses were run combining the data from the MDRS study with the performance data from the recognition and detection studies. Multiple regressions were run, with the performance data from the recognition and detection studies serving as dependent variables (d, beta, percent correct, and reaction time) and the scores from the six subscales plus a total score from the MDRS providing the predictor or independent

variables. Multiple regressions were run for each dependent variable against subscale scores (predictors) for each display type and when collapsed across display type.

Discussion. The results of the multiple regressions demonstrated that statistically significant relationships do exist between the performance measures of the tachistoscopic method and the MDRS. The MDRS can reliably predict between 11 and 67% of the variablility in the t-scope measures of performance. Thus, the two methods do converge.

When the MDRS subscales were considered in isolation, collapsing across the dependent measures, and display type, it was found that FO (format) and CD (content density) each appeared nine times in the multiple regressions, indicating that these subscales are most critical in predicting performance. CF (cognitive fidelity) and CI (content integration) were the second most frequent and therefore salient in predicting performance, each occurring five times. GA (general acceptance) and sum (the total instrument score) appeared least frequently, (four and two times respectively). It is important to note that all six subscale scores and the total score were critical in prediction for the various multiple regressions. The researchers would also expect the critical subscales to change, dependent on display type and performance measure.

Three other findings of interest resulted from the multiple regression analysis: the dependent measures of d, Beta, and percent correct from the detection study were negatively correlated with the MDRS; the dependent measures of d, Beta, percent correct, and response time from the recognition study were positively correlated with the MDRS subscale scores and total score. The detection study only produced significant R^2 results with the star display, and the recognition study produced significant multiple regression results with the bar and meter displays.

To understand these results, it is necessary to consider the methodologies used in the detection and recognition studies, and of course the displays themselves. The detection study methodology sought to discriminate between displays based solely on abnormal parameter

perception; a low level phenomena when considering human memory and learning. On the other hand, the recognition study sought to discriminate between displays based on abnormal parameter recognition; a much higher-level cognitive process. Also, since the MDRS asked the subject to rate the displays on content density, integration of content, organization of format, clarity, ease of processing, and aid to decision making, the subjects evidently rated the displays not based on purely perceptual aspects, but on how well, in their opinion, the display presented information for ease and accuracy of use. These postulates must be coupled with the fact that stars achieved significance for the detection study whereas bars and meters achieved significance only on the recognition study. The star display was unfamiliar to the operators and they did not have time to become so familiar with the display that they could accurately predict their performance with a recognition task using the MDRS. Thus, the relationship between ratings and actual performance is attenuated. For bars and meters, however, operators can predict performance because they have experience with these formats. The detection study collected lower-level cognitive data not being directly assessed by the MDRS and somewhat different than what the operator would normally consider in answering the general question of how well the display presents information for ease and accuracy of use. Thus the bar and meter displays did not give significant results with the detection data and the MDRS; however, since the star display was unfamiliar, the operators responded to the MDRS in a manner different than that for either bars or meters, thereby causing the purely perceptual performance data of the detection study to be predicted by the subscale scores of the MDRS.

The major conclusion is that one can predict the type of performance data yielded by the t-scope studies using the MDRS. It is also true that the t-scope adds a unique portion of explained variance not covered by the MDRS. The MDRS is sensitive to differences in operator familiarity with the display and predicts different levels of cognitive functioning commensurate with the operator's prior knowledge. Research is currently being conducted to include checklist and simulation evaluation techniques in this multimethods approach to further identify and validate possible means of display evaluation.

A Multimethod Approach to Safety Parameter Display Evaluation

W.W. Banks
H.S. Blackman
D.I. Gertman
R.J. Petersen

EG&G Idaho, Inc.

Overview

- Experimental questions
- Methods
- Results
- Conclusions

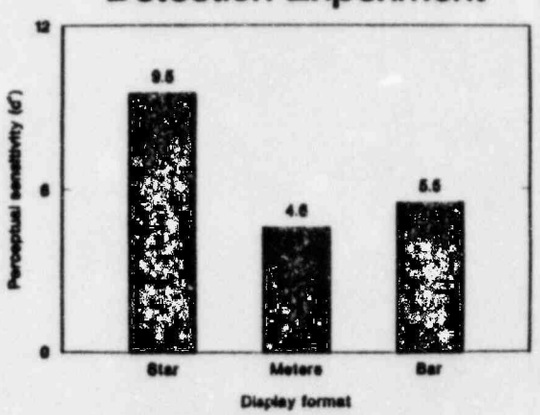
C2 0646

Experimental Questions Posed

- What is the degree of convergence between these two independent methods used in display evaluation?
- Can one of the two methods be used successfully to predict results of the other?
- What dimensions of SPD formats appear to be most crucial to operators for performance and preference?

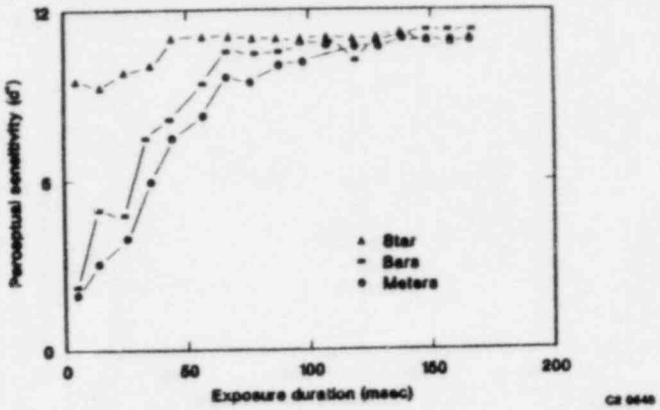
C2 0636

Detection Experiment



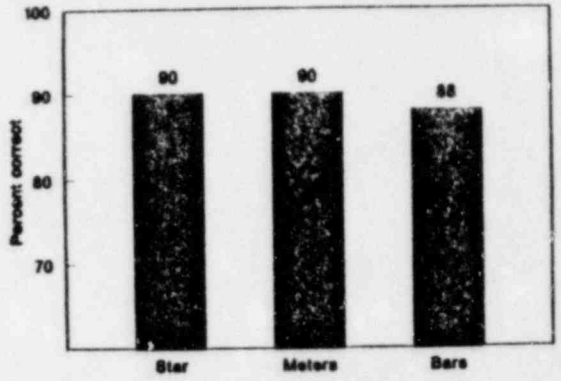
C2 0648

Detection Experiment



C2 0646

Parameter Recognition Experiment



C2 0650

Cognitive Dimensions

- Content Density (CD)
- Content Integration (CI)
- Format (FO)
- Cognitive Fidelity (CF)
- Cognitive Processing (CP)
- General Acceptance (GA)

C2 0679

1. Main effect for variable CI across display type.

df	MS	MS_{error}	F	Sig
2,51	1.675	.427	3.917	<.05

2. Main effect for variable CP across display type.

df	MS	MS_{error}	F	Sig
2,51	1.520	.407	3.728	<.05

C2 0376

Findings of Interest

- The dependent measures of d', beta, and percent correct from the detection study were negatively correlated with the MDRS.
- The dependent measures of d', beta, percent correct, and response time from the recognition study were positively correlated with the MDRS subscale scores and total score.
- The detection study results only produced significant R^2 with the star display, and the recognition study results produced significant multiple regressions with the bar and meter displays.

C2 0643

Major Conclusions

- One can predict t-scope performance data using the MDRS
- MDRS is sensitive to differences in operator familiarity with displays
- Further work utilizing checklist and simulation techniques is warranted

C2 0641

Multiple Regression Analyses

- For each dependent variable against subscale scores in four cases.
  1. Collapsed across display type
  2. Bar displays only
  3. Meter displays only
  4. Star displays only

C2 0644

Statistically Significant R^2s

d'

	Star	Bar	Meter	Collapsed
Detection	0.44			0.11
Recognition		0.47	0.59	

B

	Star	Bar	Meter	Collapsed
Detection	-0.23			
Recognition			0.65	

C2 0646

Statistically Significant R^2s (cont'd)

Percent Correct

	Star	Bar	Meter	Collapsed
Detection	0.56			0.14
Recognition		0.67	0.44	

Reaction Time

	Star	Bar	Meter	Collapsed
Detection				
Recognition		0.50		0.16

C2 0647

Major Results

- Statistically significant relationships exist
- Measures are reliable
- Can predict up to 67% of the variability
- The tachistoscopic and MDRS methods converge
- Most salient dimensions in order of importance are
  - FO (format) and CD (content density)
  - CF (cognitive fidelity) and CI (content integration)
  - GA (general acceptance) and sum (total instrument score)

C2 0643