AIIM

Association for Information and Image Management

1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910
301/587-8202

Centimeter

Inches

MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.

1 of 1

Engineering Physics and Mathematics Division

EXPERIMENTAL VALIDATION OF NAVIGATION WORKLOAD METRICS*

Jack C. Schryver

Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, TN 37831-6360

Jerry A. Wachtel

U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research/Human Factors Branch
MS NL/N-316
5640 Nicholson Lane
Rockville, Maryland 20852

"The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-84OR21400. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes."

To be Presented at the 38th Annual Meeting of the Human Factors and Ergonomics Society, October 24-28, 1994, Nashville, TN

MASTER

* Research sponsored by the U. S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, under U. S. Department of Energy Interagency Agreement No. 1886-8085-2B with Martin Marietta Energy Systems, Inc. under contract No. DE-AC05-84OR21400 with the U. S. Department of Energy.

Abstract. Advanced digital computer display interfaces in the control room may increase operator workload. Workstation monitors provide limited display area, and information is represented in large-scale display networks. Display navigation may generate disorienting effects, require additional resources for window management, and increase memory and data integration requirements. Six ORNL employees participated in an experiment to validate proposed metrics of navigation workload in the advanced control room. The task environment was a display network consisting of 25 windows resembling a simplified Safety Parameter Display System for Pressurized Water Reactors. A repeated measures design with 3 within-subjects factors was employed. The factors were task difficulty (2), navigation distance level (5), and a blocking factor (5). Participants were asked to monitor a single parameter (easy task) or two parameters (difficult task). Fourteen candidate metrics were tested. Analysis of variance of the modified task load index (MTLX) and rating subscales (confidence, disorientation, effort) demonstrated substantial support for the claim that navigation of large-scale display networks can impose additional mental load. Primary and secondary task performance measures exhibited ceiling effects. Memory probes for these tasks were inadequate because they were recognition-based and coarse. Eye gaze measures were not validated, indicating a need for more refined data reduction algorithms. Strong positive correlations were found between MTLX and both navigation duration and standard deviation of pupil diameter. Further study and increased statistical power are required to validate objective navigation workload metrics.

Experimental Validation of Navigation Workload Metrics

Jack C. Schryver
Oak Ridge National Laboratory

Jerry A. Wachtel
U.S. Nuclear Regulatory Commission

INTRODUCTION

Advanced computer display interfaces in the control room may impose increased workload as compared to traditional analog interfaces. Workstation monitors provide limited display area, while information is represented in a large-scale display network. This condition engenders the keyhole effect, where only a small fraction of the plant information base is observable at each moment. The keyhole effect has broad implications for information processing. Display navigation may generate disorienting effects, aggravate getting lost, create tunnel vision, require additional resources for window management, and increase memory and data integration requirements (Elm and Woods, 1985). The operator is faced with the task of integrating disparate information representing many complex processes from multiple displays. When a new display is called-up, an operator must search, extract, interpret, and integrate items of interest. Resource competition and maintenance of working memory schema demand cognitive resources, which generate *navigation workload*.

Workload metric validation studies have been conducted mainly in environments which differ in significant aspects from the nuclear power plant control room (NPPCR). Moreover, the term workload, in the present context, is being rather narrowly defined as the mental cost or difficulty associated with navigating large-scale display networks. An experiment was designed to validate navigation workload metrics using the sensitivity criterion in a NPPCR context.

METHOD

Participants. Six employees participated in the experiment at Oak Ridge National Laboratory (ORNL). Two volunteers were operators or supervisors at operational facilities at ORNL. Complete data was obtained for four participants. The remaining two participants completed the test trials, but no usable eye-gaze data was recorded.

Apparatus. The experiment control application ran on a Sun workstation. Eye movements were captured with a CCTV camera, which was mounted just beneath the monitor. An eye-gaze data stream was delivered over the serial port to the workstation application, which stored eye position data to disk. The application computed a randomized presentation order, controlled presentation of each trial, and captured timing, navigation behavior, and subject responses following each trial.

The eye-gaze system is an unobtrusive device that uses a pupil illumination/corneal reflection method to compute point-of-regard to a calibrated display at a rate of 30 Hz. The main components are: (1) 80386-based microcomputer; (2) video frame grabber board; (3) high-speed infrared sensitive camera and 75 mm lens containing an LED infrared source; and (4) eye-image video monitor. Proprietary software processes the eye-image using pattern recognition algorithms, and geometric calculations recover eye-gaze point of regard. Tolerance to head motion is slight; therefore, a chin rest was required to eliminate significant head motion.

Task Environment. The task environment was a computer display network consisting of 25 windows resembling a simplified Safety Parameter Display System for Pressurized Water Reactors. Each display in the network contained a simple schematic diagram, and from two to four digital displays marked by distinctively colored icons. A total of 66 parameters were represented in the 25 displays. A new set of parameter values was presented on each block of ten trials in order to prevent system state learning from easing primary task demand.

Navigation Distance. Seidler and Wickens (1992) defined several types of distance applicable to information access. *Navigation distance* is the number of steps required to arrive at a target display. A step is defined as a decision followed by a keypress, where each type of step has equal value. One problem with this definition of navigation distance is that it does not characterize decisions underlying navigation that are not associated with overt responding. Information theory was employed to define a unit of navigation distance as a single binary decision, or bit. Each bit represents a choice available to a user navigating the network. The first decision the user makes is to either move up or down the hierarchy, remain at the present display, or go directly to the top level. The four choices are equivalent to two bits. If the decision is to go to the top or stay put, further decision-making is not required, and the total navigational distance is 2. A decision to move up may present the user with an option to shift either one, two, or three levels, or 1.585 bits. Thus, total navigation distance for the step is 2 plus 1.585, or 3.585.

Design. A repeated measures design with 3 within-subjects factors was employed. The factors were task difficulty (2), navigation distance level (5), and a blocking factor (5). Participants were asked to monitor and remember a single parameter (easy task) or two parameters (difficult task). Trials were presented in six blocks, where each block contained one set of all combinations of navigational distance and task difficulty. The first block was used for navigation practice.

Fourteen candidate metrics were tested: primary task performance, secondary task performance, navigation duration, route efficiency, number of displays, a modified NASA task load index (MTLX) with three subscales, and five eye gaze measures. Two subscales from the NASA-TLX were imported into the MTLX: effort and performance. A third subscale, frustration, was modified to detect disorientation effects produced by navigation. The MTLX was defined as the simple mean of the three rating scores (Nygren 1991).

Eye-gaze metrics have a long history in workload research. Beatty (1982) summarized a number of studies which reported pupil diameter sensitivity to demands of diverse cognitive activity, including mental arithmetic, short-term memory, language processing, perceptual detection, sustained attention and selective attention. Blink rate has been found to decrease as visual demand is increased in the flight environment. A blink was defined in this study as an on-screen loss-of-tracking from 33 to 333 msec. in duration. A simple measure of eye gaze complexity was adopted. Planar deviation between successive samples was computed whenever there was no loss-of-tracking. Planar deviation is simply the Euclidean distance between pixel locations on successive samples. The mean and standard deviation of planar deviation and pupil diameter were computed.

Procedure. An unbiased algorithm generated a randomized presentation order by taking a semi-random walk through the display space. Each trial began with a window matching the target window from the previous trial. A pop-up task window presented the primary and secondary task. The window displayed a request to monitor either one (easy task) or two (difficult task) randomly-selected parameters. The identity of the target window (determined by the randomized presentation order) and a three-letter request code were also displayed. The task window closed after six seconds. Instructions stressed the importance of the monitoring task as compared to remembering the request code.

The participant monitored parameter(s) on the underlying window, and navigated toward the target window by clicking the mouse over appropriate icons. The exact number of intervening windows for a given trial was determined by the navigation distance level selected for that trial. The participant was urged to proceed as fast as possible to the target window without making mistakes. When the target window was reached, the DONE box was clicked. This action terminated timing and eye gaze collection, and opened the pop-up response window shown in Figure 1. If the target window was not in view, however, when the DONE box was marked, an error window popped up instead, advising the user to proceed to the correct target window. The window remained on-screen for six seconds before automatically closing.

The first question at the response window posed a three-alternative multiple choice requiring the participant to select the correct interval for a monitored value. If two variables were monitored, the application randomly selected one for testing. The correct answer was randomly assigned to the first, second, or third choice. The second question was a six-alternative multiple choice probe for the request code. Five distractor items were randomly selected from a 100-word list. The third question prompted the participant to provide subscale ratings. All responses were indicated by clicking in the appropriate box.

Logsheet	
<p>1. The value of Steam Supply Rate is:</p> <div style="display: flex; align-items: flex-start;"> <div style="margin-right: 20px;"> ☐ less than 1903 ☐ between 1903 and 2127 ☐ greater than 2127 </div> <div style="border: 1px solid black; padding: 10px; text-align: center; width: 60px;">OK</div> </div> <p>2. The request code is:</p> <div style="display: flex; justify-content: space-between;"> <div> ☐ MAY ☐ BEE ☐ GEM </div> <div> ☐ MOP ☐ WOE ☐ SKY </div> </div> <p>3. Rate the following on a 10-point scale:</p> <p>effort: 12345678910</p> <p>confidence in answer to (1): 12345678910</p> <p>disorientation or frustration: 12345678910</p>	

Figure 1. Pop up response window.

Participants were asked to guess if they did not know the correct answers. Finally, clicking in the OK box closed the response window and brought up a new trial. If the primary task response was incorrect, a window appeared to identify the error.

RESULTS

Fourteen separate analyses of variance (ANOVA) were conducted so that each candidate workload metric could be independently validated. The primary validation criteria in terms of ANOVA were significant effects of navigation distance, task difficulty, and their interaction. A secondary criterion was that practice should reduce measured load.

Primary task performance and secondary task performance were not validated on the three main criteria. However, secondary task performance did register a practice effect. Both dependent measures exhibited a ceiling effect in which maximum performance was nearly achieved for most conditions. One explanation of these findings is that the primary task was easy enough to guarantee uniformly high performance, leaving sufficient residual capacity for the secondary task. Other possibilities were that the secondary task drew upon a separate resource pool, or that probe items did not adequately test the memory trace available at response time.

Route efficiency was an increasing function of navigation distance. One would not expect navigation to become more efficient at longer distances. Navigation duration was validated on one of the three primary criteria (navigation distance),

and also on the secondary criterion (practice). The navigation distance effect could almost be predicted by definition, however.

The three rating subscales (confidence, disorientation, effort) and the MTLX were validated on two of three main criteria including navigation distance (Figure 2). The confidence rating and MTLX experienced significant effects of practice (block). These results represent strong evidence for validation, considering the modest number of test participants.

Figure 2 Navigation distance effect on MTLX.

The ANOVAs did not uncover a single significant effect in support of main or secondary criteria for blink rate, mean pupil diameter, standard deviation of pupil diameter (SDPD), planar deviation, or standard deviation of planar deviation. A marginal effect of navigation distance suggested that pupil diameter increased with navigation distance, which is consistent with the notion that dilation accompanies greater workload demand.

Correlation analyses offered converging evidence for subjective measures in cases for which subjective and objective measures were strongly associated. Strong positive correlations between MTLX and both navigation duration and SDPD (Figure 3) provided cross-validation for MTLX. Each subscale appeared to contribute to the omnibus score, and had strong correlations with either navigation duration or SDPD. Disorientation appeared to be the most sensitive component of MTLX.

Figure 3 Regression of MTLX on SDPD.

DISCUSSION

Analysis of the MTLX demonstrated substantial support for the claim that navigation of large-scale display networks can impose additional mental load. Further study and increased statistical power is required to complete validation of objective workload metrics, but the evidence suggests that navigation workload could be a source of risk in the advanced control room.

Procedures for primary and secondary task testing should be modified for regulatory review of navigation workload in the control room. Memory probes used in the present study were inadequate because they were recognition-based and coarse. More sensitive measures would determine relative strengths of memory traces.

Refinements are needed for calculation of eye gaze metrics. Data reduction algorithms should: (1) logically separate blink rate and planar deviation; (2) exclude saccades from pupil diameter calculations; and (3) eliminate bias from planar deviation resulting from exclusion of loss-of-tracking episodes.

REFERENCES

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276-292.

Elm, W. C., and Woods, D. D. (1985). Getting lost: A case study in interface design. pp 927-931 in *Proceedings of the Human Factors Society 29th Annual Meeting-1985*.

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors*, 33(1), 17-33.

Seidler, K. S., and Wickens, C. D. (1992). Distance and organization in multifunction displays. *Human Factors*, 34(5), 555-569.

DATE

FILMED

9 / 7 / 94

END