

Evaluating User Interfaces with CWS: Competitive Usability Testing of Monitor Sizes

J. Shawn Farris*, John Raacke, & James Shanteau
 Department of Psychology, Kansas State University

The CWS expert performance index was applied as a tool for evaluating computer usability in CTEAM (Controller Teamwork Evaluation and Assessment Methodology), an air traffic control microworld environment. CWS integrates discrimination and consistency such that larger CWS scores indicate better performance. Seven experienced participants routed aircraft to assigned destinations in the CTEAM task environment with 3 different computer monitor sizes (17, 19, and 21 in.) and 2 levels of aircraft density (high and low). As expected, aircraft density affected performance such that scenarios that were more complex led to lower CWS scores. In addition, monitor size influenced CWS scores, which is consistent with previous applied research. The potential for using CWS in human-computer interaction research is discussed.

Introduction

Human-Computer Systems

Computers have become ubiquitous in our day-to-day lives because of, among other reasons, their ability to increase user productivity and performance. A crucial component in the design lifecycle of these computer technologies is usability testing. Usability testing is a set of methodologies aimed at evaluating the ease-of-use and/or productivity and performance gains with computer systems for the products' target population. Competitive usability testing is one such method that can be used to evaluate usability differences between similar designs of a product being developed or between a new design and a competitor's product (Schneiderman, 1998). However, one problem with competitive usability testing is deciding what measures are valid for the comparison of competing interfaces.

If one assumes that computers and humans form a system for completing tasks, then it is logical to assume that a measure of human expert performance would be an appropriate measure of human-computer system performance. That is, a computer that better supports the human's expertise will increase the performance and productivity of that human-computer system. Based on this logic, it is possible that a measure of expert performance (i.e., CWS) could be used as a measure to evaluate the usability of computer systems.

CWS to Evaluate Human-Computer Systems

The Cochran-Weiss-Shanteau (CWS) index of expert performance was tested to determine whether it could be used as a measure for competitive usability testing. CWS integrates two necessary conditions for expert skill. The first is consistency, as argued by Einhorn (1972, 1974). Experts must make reliable judgments of identical stimuli; unreliable judgments serve as evidence against expertise. The second necessary condition is discrimination ability (Hammond, 1996), i.e., experts should be able to differentiate stimuli based on subtle differences that non-experts are typically insensitive to. CWS integrates these two conditions by taking the ratio of discrimination to inconsistency, such that higher CWS scores are more indicative of expert performance. That is, experts should be consistent in their discriminations of stimuli in their domain.

CWS has been successfully applied as a performance measure to several pre-existing datasets, three of which are presented in Shanteau, Weiss, Thomas, and Pounds (2002): auditing (Ettenson, 1984), personnel hiring (Nagy, 1981), and livestock judging (Phelps & Shanteau, 1978). In addition, it has been successfully used to track performance and skill development in a simulated air-traffic control environment (Friel, Thomas, Raacke, & Shanteau, 2001), the Controller Teamwork Evaluation and Assessment Methodology (CTEAM) microworld (Bailey, Broach, Thompson, & Enos, 1999).

To date, CWS has only been used to evaluate the human in the human-computer system. Therefore, the reported experiment explored the possibility that CWS could be used as a competitive usability testing measure. To do this, a monitor size manipulation was introduced because, based on applied and theoretical research, there should be a performance change with changes in monitor size. If so, then CWS should be sensitive to these changes.

Monitor Size

What is the optimum computer display size? Simmons and Manahan (1999; Simmons, 2001) demonstrated that larger monitor sizes have a positive effect on the performance time of tasks that involve the searching of large amounts of data, but not on other tasks (the "search vs. create dichotomy"). Based on this research, if we consider an air-traffic control simulation to be analogous to searching through large amounts of data (e.g., visual search for aircraft and potential problems), then a larger monitor should yield better performance.

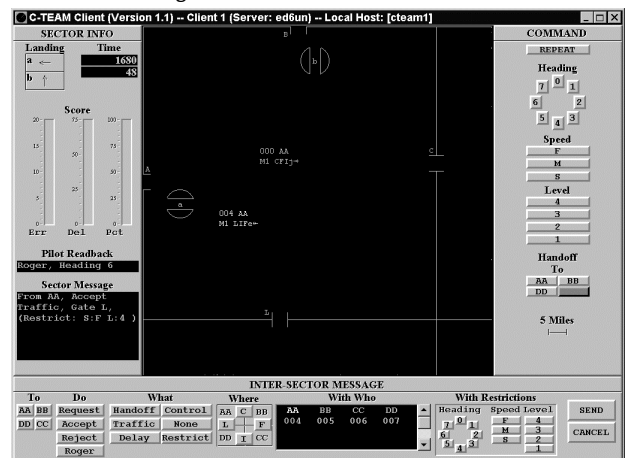
Method

CTEAM Environment

In order to examine computer usability, the CTEAM computer-simulated microworld was used as a low fidelity simulation of Air Traffic Control (ATC). Developed and maintained by the Federal Aviation Administration (FAA), CTEAM is a simplified version of a real ATC task that allows for individual or team operator designs. The reported study used the single-sector (or individual) version of CTEAM.

CTEAM operators control traffic in their virtual airspace by using a command toolbar, located in the right hand corner of the screen (see Figure 1), to issue heading, speed, and altitude commands. Aircraft are either landed at an airport in the operators sector or sent through an adjacent exit gate.

Figure 1. CTEAM Environment



* farris@ksu.edu; Department of Psychology, Kansas State University

CTEAM records various performance measures, including operational errors, percentage of aircraft reaching their destination, and time through sector. Time through sector (TTS; the time from when an aircraft appears in an operator's sector to the time it reaches its final destination) was the performance measure used in the reported study.

Design & Procedure

In a single-sector version of CTEAM, participants completed high and low aircraft density scenarios using 17, 19, and 21-inch monitors in a within-subjects design. The low-density scenario contained 12 aircraft, whereas the high-density scenario contained 36 aircraft. In order to compare scores across Aircraft Density and Monitor Size, the 12 aircraft from the low-density scenario were embedded within the high-density scenario. These 12 aircraft were used for all analyses.

All scenarios were presented on Sony Trinitron FD monitors connected to NCR-3230 486 computers. All monitors were set 123 cm high and 15 cm from the edge of the desk. Participants directed aircraft to exit gates or airports using Kensington Expert Mouse trackballs.

Seven experienced participants completed two repetitions of the same scenario in each 1.5-hour session. After an initial practice session, participants completed five sessions per week (one for each day) during two and a half weeks, yielding 13 sessions.

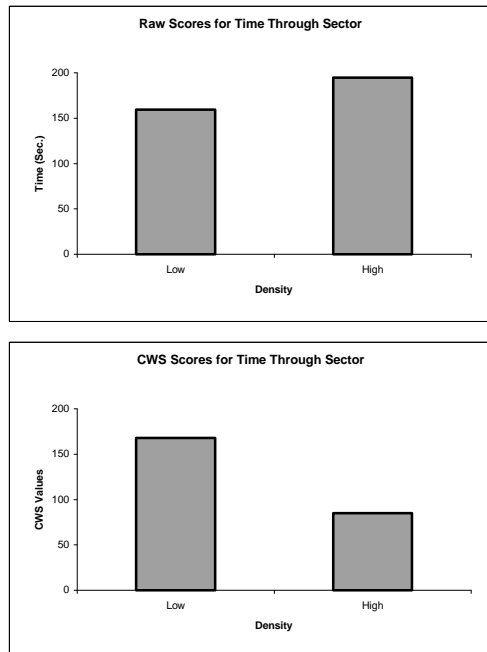
Results & Discussion

CWS analysis was performed on the dependent variable TTS. The results below are broken into two sections: Aircraft Density and Monitor Size.

Aircraft Density

Results indicated that as aircraft density increased, raw TTS also increased (i.e., aircraft took longer to exit the scenario as aircraft density increased). Accordingly, CWS analysis showed that as aircraft density increased, CWS scores declined (i.e., expert performance decreased as aircraft density increased; see Figure 2). These results are consistent with Friel, Thomas, Raacke, & Shanteau's (2001) findings. Interestingly, CWS appears to be more sensitive to changes in aircraft density than raw time.

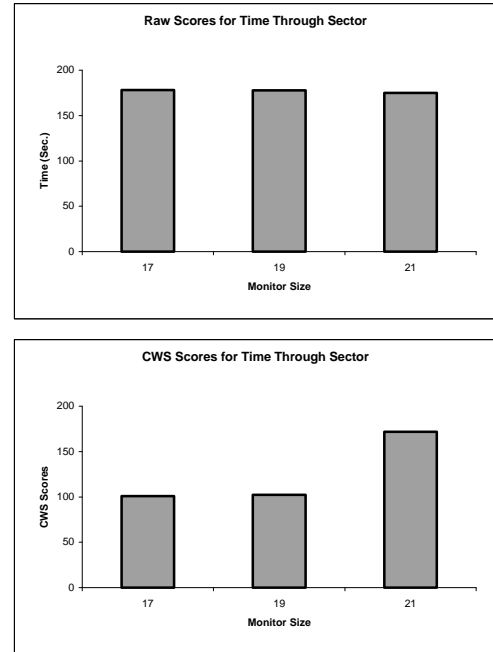
Figure 2. Aircraft Density



Monitor Size

As can be seen in Figure 3, the largest monitor yielded the fastest TTS (i.e., aircraft exited the scenario faster as monitor size increased). Accordingly, as monitor size increased, CWS scores also increased (i.e., expert performance increased as monitor size did; see Figure 3). These results are consistent with those obtained by Simmons and Manahan (1999; Simmons, 2001). Again, CWS appears to be more sensitive to changes in monitor size than raw time and, therefore might be an effective tool for competitive usability testing.

Figure 3. Monitor Size



References

Bailey, L. L., Broach, D. M., Thompson, R. C., & Enos, R. J. (1999). *Controller teamwork evaluation and assessment methodology: (CTEAM): A scenario calibration study.* (DOT/FAA/AAM-99/24). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161.

Einhorn, H. J. (1972). Expert measurement and mechanical combination. *Organizational Behavior and Human Performance*, 7, 86-106.

Einhorn, H. J. (1974). Expert judgment: Some necessary conditions and an example. *Journal of Applied Psychology*, 59, 562-571.

Ettenson, R. (1984). *A schematic approach to the examination of the search for and use of information in expert decision making.* Unpublished doctoral dissertation, Kansas State University.

Friel, B. M., Thomas, R. P., Raacke, J., & Shanteau, J. (2001). Utilizing CWS to track the longitudinal development of expertise. *Proceedings from the 45th Annual Meeting of the Human Factors and Ergonomics Society*. Minneapolis/St. Paul, MN.

Hammond, K. R. (1996). *Human judgment and social policy.* New York: Oxford University Press.

Nagy, R. H. (1977). *How are personnel selection decisions made? An analysis of decision strategies in a simulated personnel selection.* Unpublished doctoral dissertation, Kansas State University.

Phelps, R. H., & Shanteau, J. (1978). Livestock judges: How much information can an expert use? *Organizational Behavior and Human Performance*, 21, 209-219.

Schneiderman, B. (1998). *Designing the user interface, 3rd edition.* Reading MA: Addison-Wesley.

Shanteau, J., Weiss, D. J., Thomas, R., & Pounds, J. (2002). Performance-based assessment of expertise: How can you tell if someone is an expert? *European Journal of Operations Research*, 136, 253-263.

Simons, T. (2002). What's the optimum computer display size? *Ergonomics in Design*, 9 (4), 19-25.

Simmons, T., & Manahan, M. (1999). The effects of monitor size on user performance and preference. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting*, 1393.