Menu Design Based on Expert Knowledge Structures: A Validation

Christine M. Covas-Smith[1, 2], Kenneth D. Jackson[1, 2], Russell J. Branaghan[1, 2], & Craig Eidman[1]
Air Force Research Laboratory[1], Arizona State University[2]

This project evaluated the effectiveness gained by redesigning a user interface based on the knowledge structure of domain experts. Scenario-based simulation enables realistic training in synthetic environments. As the instructor remains essential to facilitating effective scenario-based training, their user interface should be efficient. Previous research modeled knowledge structures used to redesign a menu within a scenario generation interface. The current research evaluated the redesigned menu based on the ascertained knowledge structures. To evaluate the redesigned menu, participants were timed as they selected target items in the original and redesigned menus. Results indicated that experienced users selected items in the redesigned menu significantly faster, and preferred the redesigned menu to the original. Thus, this result demonstrates tangible benefits to organizing menus according to expert knowledge, even when users are familiar with a menu structure. Using methods that follow the general principles presented here will enable designers to uncover expert knowledge and efficiently configure user interfaces, thereby improving training effectiveness.

INTRODUCTION

Simulators are utilized frequently (Andrews & Bell, 2000; Harris & Kahn, 2003; Tannenbaum & Yuki, 1992) to engage users in synthetic environments in medicine (Park et al., 2010; Schwid et al., 2002), aviation (Wickens, Helleberg, & Xu, 2002), military (Andrews & Bell, 2009), space exploration (Cyril, Jaar, & St-Pierre, 2000), driving (Fisher, Laurie, & Glaser, 2002), and other fields. Through simulation, trainees experience realistic decision environments. Furthermore, debriefing capabilities enable students to rerun simulations, consider alternative actions, and discuss training episodes with an instructor. This capability to experience and revisit training scenarios cannot be replicated with audio-visual presentation, lecture, or paper and pencil methods.

Military pilots frequently engage in scenario-based simulations (SBS) while training (Andrews & Bell, 2009; Bell & Waag, 1998; Fowlkes, Dwyer, Oser, & Salas, 1998). The scenario acts as the curriculum by providing situations that expose the trainee to the knowledge and skills to be learned. Efficient training episodes result from well planned scenarios that ensure trainees practice the correct content and receive appropriate and timely assessment and feedback.

The eXpert Common Immersive Theater Environment (XCITE) is computer based simulation software used at the Air Force Research Laboratory to build, manage, and execute scenarios for pilot and other military training. XCITE provides physics-based aircraft and threat models, and the ability to control many entities within a scenario concurrently, facilitating realistic complexity and diversity within a task (Eidman, Crane, Kam, Gehr, & Zamba, 2007). Using XCITE, military trainees perform simulated missions replicating the important aspects of their task, including equipment, team, terrain, objective, and enemy.

Scenario building and presentation are accomplished using an instructor operator station (IOS) (Ramesh & Sylla, 1990). The IOS should enable instructors to perform their functions efficiently. Typically, instructors are not computer programmers, so the user interface design of the IOS is crucial to instructor effectiveness. Unfortunately, despite its technical sophistication, XCITE's IOS user interface was designed with minimal input from instructors.

It has been shown that menu organization is enhanced by user input (Hayhoe, 1990; Roske-Hofstrand & Paap, 1986). Previous research by Covas, Jackson, Branaghan, and Eidman (2010) redesigned one menu structure in the XCITE interface (the "declutter" menu; as depicted in Figure 1) using knowledge elicitation via card sorting and representation as well as hierarchical cluster analysis (HCA). This approach to menu redesign was inspired by a cognitive engineering premise that user interfaces should often mirror the knowledge of expert users (McDonald, Dayton, & McDonald, 1988; McDonald & Schvaneveldt, 1988; Paap & Cooke, 1997; Roske-Hofstrand & Paap, 1986), as well as the principle of proximity compatibility (Wickens & Carswell, 1995).

This paper describes activities to validate the redesign using a reaction time paradigm and preference selection. Specifically, we measured the speed with which experienced XCITE users (all of whom were expert instructors) could locate and select menu items in the new and the old menus. Additionally, we investigated which of the two menus these experienced XCITE users preferred.

If the new menu structure is superior to that of the old interface, and if the methods of card sorting and HCA were valid, we would expect participants to find and select items faster from the redesigned menu. Further, we would expect them to prefer the redesigned menu to the original. On the other hand, since the participants were all current and frequent users of the original XCITE menu structure, we might expect that they will locate and select items faster using the original menu structure, and they might prefer it. Because of the participants' familiarity with the original menu, this represents a strong test of the redesign. Further, because XCITE is often used for operational training that is not flight-centric, we included XCITE users with a variety of military operational experience, not just pilots.

Participants were presented with menu-item labels to locate in both the original and redesigned menus, and were timed as they located, and clicked on them. After completing the reaction time tasks, participants were shown both menu

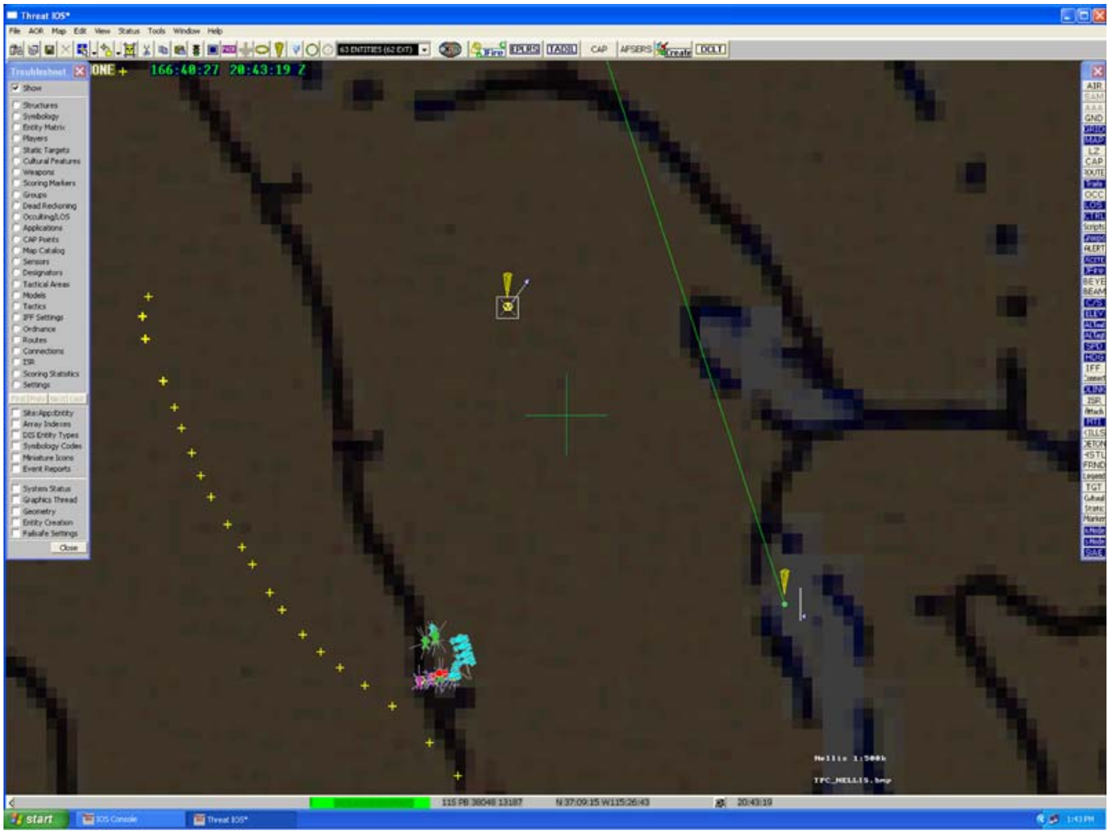

Figure 1. Screen capture of XCITE's graphical user interface. The "declutter" menu is shown on the right hand side.

configurations and asked to indicate which structure they preferred. We predicted that participants would locate and select items on the redesigned structure faster than the original menu structure and that participants would prefer the redesigned menu structure.

METHOD

Participants

Thirteen XCITE users (12 males and one female) participated in the experiment. All participants received and provided informed consent. Mean age of participants was 41 (S.D. = 9.9, range = 27-50 years). Four of the participants had used XCITE for 6 months or less, 2 had used XCITE between 7 and 12 months, 4 between 1 and 3 years, and 3 for more than 3 years. Three participants indicated that they used XCITE less than once per year, 3 indicated that they used it monthly, 3 weekly, and 2 daily. Five of the participants had operational experience as a military pilot, and 5 had operational experience as a Joint Terminal Attack Controller, one was an Intelligence Officer, one was Army Infantry, and one had no operational experience, but wrote instructional scripts used in XCITE training.

Stimuli and apparatus

The stimuli consisted of the original and redesigned "declutter" menus. The original "declutter menu" is depicted in Figure 1 (far right hand side). Table 1 depicts the original order of the declutter menu (note that the items were originally in a vertical column in the interface). The order of the items in Table 2 represents the product of a previous study (Covas et al., 2010) where card sorting was used to elicit the relationships among concepts in the "declutter" menu. We then used these relationships along with HCA to determine that the information should be ideally represented using six clusters.

The experimental trials were administered via Inquisit (Inquisit, 2006), a data collection software package used to design surveys and reaction time experiments. Inquisit enabled the experimenter to send a link directing participants to an experiment, and then download a java applet to the participant's computer, enabling the computer's internal clock to be used for collecting reaction times.

Procedure

Prior to the evaluation, participants were emailed an informed consent document. Upon return of the document, they were sent a link to the software. On opening the link to the experiment, participants were asked to complete a brief demographic survey that included questions related to XCITE and operational experience.

After completion of the demographic survey, participants were provided with task completion instructions. For the menu-search task, participants were shown the name of an item taken from the "declutter" menu, in large print on the center of their computer screen. As items were displayed, participants were asked to locate and click the item within the menu on the right hand side of the screen. Menu order was counterbalanced across participants. After completing the task with one menu, participants were then provided with the alternate menu structure. There were 44 items in both the current menu and redesigned "declutter" menus. After all trials were completed, participants were asked to select which of the two menu structured they would prefer to use. Thus, participants performed a total of 88 individual trials all of which took less than 30 minutes.

Table 1. Order of Items in Original "Declutter" Menu

Air	Trails	Bullseye	Connect	Legend
SAM	OCC	Beam	DLINK	Target
AAA	LOS	C/S	ISR	Cultural
Ground	Control	Elevation	Attach	Static
Grid	Scripts	Alt-msl	MTI	Marker
Map	Groups	Alt-agl	Kills	K-Mode
LZ	Alert	Speed	Detonate	S_mode
CAP	XCITE	Heading	Hostile	S.A.E.
Route	J-Fire	IFF	Friendly	

Note: To preserve space in the document, the menu is depicted as a table in 5 columns. In the actual interface, the menu is simply a one-column list of the forty-four items.

Table 2. Order of Items of the Redesigned Menu in the Six Clusters.

Beam	Detonate	AAA	Bullseye	Alt-agl	Alert
DLINK	Friendly	Air	C/S	Alt-msl	Control
Groups	Hostile	Attach	K-Mode	Cap	Cultural
IFF	Kills	Connect	OCC	Heading	Elevation
LZ	Trails	Ground	S_mode	Route	Grid
Marker		SAM	S.A.E.	Speed	ISR
MTI		Scripts		Static	JFIRE
		XCITE		Target	LOS
					Map

Note: Several items within the two menu structures were given slightly different labels than in the original menu. The terms were labeled as either likely difficult for novice users (GND, CTRL, BEYE, CLS, ELEV, ALTmsl, ALTagl, SPD, HDG, DETON, HSTL, FRND, TGT, SMODE) or were unable to be programmed as originally labeled in XCITE (K-MODE, SMODE, S.A.E., CLS).

Results

As hypothesized, participants found and selected items in the redesigned menu (*M* = 3,903 ms, *SD* = 672.9) faster than in the original menu (*M* = 4,594 ms, *SD* = 1067.8; *t* (12) = 2.47, *p*=0.03; Cohen's d = 0.68). Overall, subjects responded 15% faster with the menu designed based on the expert knowledge structures. Additionally, participants preferred the redesigned menu structure to the original structure (χ^2 (1) = 6.23, *p* = 0.013).

Overall, we found that the redesigned menu structure significantly improved reaction and search time to locate items within the menu. Further, not only were the interactions with the menu improved by the redesign, we also found that even the majority of experienced users preferred the redesigned menu structure. These results support the benefits of using expert knowledge as a starting point for the design of menu structures.

DISCUSSION

Scenario based simulation software, such as XCITE, is widely utilized in military training systems thus enabling trainees to practice tasks in a realistic context. Although a simulator is used, instructors will remain essential as they script the scenarios, control entities and provide feedback. As a result, the instructor-operator station user interface should be usable to prevent cognitive overload. Unfortunately, the usability of these products frequently takes a back seat to functionality because programmers with little training in human factors are responsible for the user-interface design.

A cognitive engineering approach (McDonald & Schvaneveldt, 1988; Norman, 1986, 1988) to user interface design exploits the knowledge structures of domain experts. In this way, concepts that are close together in the domain expert's knowledge structure are placed close together in the user interface, thus adhering to the principle of proximity compatibility (Wickens & Carswell, 1995). Since the instructors in the domain served by XCITE are typically experts who possess many years of operational experience, XCITE represents a prime candidate for this approach.

Previous research (Covas et al., 2010) employed hierarchical card sorting to elicit the relationships among concepts in XCITE's largest (44 items) and most frequently used menu, the "declutter" menu. This paper reported on activities to validate a redesigned menu structure in the XCITE IOS using a reaction time paradigm and preference selection.

In the current study, XCITE users were shown target menu items, and asked to locate these items in the original or redesigned menu as quickly as possible. Participants were significantly faster at locating items in the redesigned menu than in the original menu. Moreover, participants preferred the new menu structure over the original. These findings are especially strong since all of the participants were active users of the original XCITE menu system.

As demonstrated in previous research, (Hayhoe, 1990; McDonald & Schvaneveldt, 1988; Roske-Hofstrand & Paap, 1986), the findings from both the current study and the previous study (Covas et al., 2010) provide support for organizing menu systems based on user knowledge. Designers, engineers and programmers are likely experts in design and engineering; however, they are rarely experts in the domain they design for. Consequently, methods like the one demonstrated here, that enable designers to uncover expert knowledge structures and then to configure the interface accordingly, would be advantageous. This seems particularly helpful for situations where designers and engineers need to design and redesign rapidly.

Importantly, the current study provides validation using both performance and preference data. Moreover, the participants were active users of XCITE who were quite familiar with its original menu structure and completely unfamiliar with the redesigned menu. This finding suggests that when a redesign reflects user knowledge, the benefits of changing the design can outweigh the benefits of remaining with an original design. As demonstrated here, if the original structure is bad enough, and the new design is good enough, making design changes can have a significant and immediate impact on user performance and preference. One strong point of this paradigm is that we used different participants to construct the menu and evaluate the menu. While all were "XCITE users", they had different operational backgrounds and different levels of expertise.

A limitation of this research is that it focuses on only the fastest and simplest type of menu search, identity matching (Paap & Cooke, 1997). This consists of the user searching and selecting the same item. For example, if the stimulus item was "Map", the target item was also "Map". A different type of search, equivalence matching (McDonald, Stone, & Liebelt, 1983) occurs when the user knows what he is looking for, but does not know its name. In this type of task, the user might receive a stimulus such as "a picture that shows how things are laid out in space". The target item, of course, would be "Map". This type of matching requires a more thorough semantic analysis. Further, people often have difficulty remembering the names of specific menu items. Future research should investigate this search type.

Although this research demonstrated improved speed and preference this may not translate into actual improved training effectiveness. On the other hand, one might expect the improved design to reduce the instructors' cognitive load, leaving more cognitive resources available to dedicate to their trainees and the curriculum. Furthermore, this research paradigm investigated only one method of knowledge elicitation, one method of representation, and one type of menu. Future research should focus on determining the generality of these findings to different methods of elicitation, representation, and different types of menus (e.g., icon based).

ACKNOWLEDGEMENTS

This work was supported by U.S. Air Force contract FA8650-05-D-6502 to L-3 Communications (Link Simulation and Training). We would like to thank Clinton Kam and Adam Pohl (U.S. Air Force) for providing computer and software support and Dr. Byron Pierce for assistance with earlier versions of this document.

REFERENCES

Andrews, D., & Bell, H. (2000). Simulation-based training. In S. Tobias & J. Fletcher (Eds.), *Training & retraining: A handbook for business, industry, government, and the military* (pp. 357-384). New York: American Psychological Society.

Andrews, D., & Bell, H. (2009). A virtual environment application: Distributed Mission Operations. In J. Cohn, D. Nicholson & D. Schmorrow (Eds.), *Virtual Environments for Training and Education: Developments for the Military and Beyond* (pp. 77-84). Westport, CT: Praeger Security International.

Bell, H., & Waag, W. (1998). Evaluating the effectiveness of flight simulators for training combat skills: a review. *International Journal of Aviation Psychology, 8*, 223-242.

Covas-Smith, C. M., Jackson, K. D., Branaghan, R. J. & Eidman, C. (2010). *Eliciting and Representing Knowledge Structures of Expert Pilots: Guidance for Menu Design*. Manuscript in preparation.

Cyril, X., Jaar, G., & St-Pierre, J. (2000). *Advanced space robotics simulation for training and operations.* Paper presented at the AIAA Modeling and Simulation Technologies Conference.

Eidman, C., Crane, P., Kam, C., Gehr, S.E., & Zamba, M. (2007). Exercise pacific link 2: distributed training for air battle managers. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, (7104), 1-12.

Fisher, D., Laurie, N., & Glaser, R. (2002). Use of a fixed base driving simulator to evaluate the effects of experience and PC-based risk assessment training on drivers' decisions. *Human Factors, 44*(2), 287-302.

Fowlkes, J., Dwyer, D., Oser, R., & Salas, E. (1998). Event-based approach to training (EBAT). *International Journal of Aviation Psychology, 8*(3), 209-221.

Gentner, D., & Stevens, A. (1983). *Mental models*. Hillsdale, NJ: Erlbaum.

Harris, D., & Kahn, H. (2003). Response time to reject a takeoff. *Human Factors and Aerospace Safety, 3*, 165-175.

Hayhoe, D. (1990). Sorting-based menu categories. *International Journal of Man-Machine Studies, 33*, 677-705.

Inquisit. (2006). *Inquisit 2.0.60616*. Seattle, WA: Millisecond Software.

Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge, MA: Harvard University Press.

McDonald, J., Dayton, T., & McDonald, D. (1988). Adapting menu layout to tasks. *International Journal of Man-Machine Studies, 28,* 417-436.

McDonald, J., & Schvaneveldt, R. W. (1988). The application of user knowledge to interface design. In R. Guindon (Ed.), *Cognitive Science and its Application for Human Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum & Associates.

McDonald, J., Stone, J., & Liebelt, L. (1983). *Searching for items in menus: The effects of organization and type of target.* Paper presented at the Human Factors Society, Norflok, VA.

Norman, D. A. (1986). Cognitive engineering. In D. A. Norman & S. W. Draper (Eds.), *User-centered system design: New perspectives in human-computer interaction* (pp. 31-61). Hillsdale, NJ: Erlbaum.

Norman, D. A. (1988). *The Psychology of Everyday Things*. New York: Harper-Row.

Paap, K. R., & Cooke, N. J. (1997). Design of menus. In M. Helander, T. K. Landauer & P. Prabhu (Eds.), *Handbook of Human Computer Interaction*. Amsterdam: Elsevier Press.

Park, C., Rochlen, L., Yaghmour, E., Higgins, N., Bauchat, J., Wojciechowski, K., et al. (2010). Acquisition of Critical Intraoperative Event Management Skills in Novice Anesthesiology Residents by Using High-fidelity Simulation-based Training. *Anesthesiology, 112*(1), 202-211.

Ramesh, R., & Sylla, C. (1990). A framework for the optimal design of instructor / operator stations in flight simulators. *IEEE Transactions on Systems, Man, and Cybernetics, 20*(3), 571-581.

Roske-Hofstrand, R. J., & Paap, K. R. (1986). Cognitive networks as a guide to menu organization: An application in the automated cockpit. *Ergonomics, 39*, 1301-1311.

Salas, E., Wilson, K., Priest, H., & Guthrie, J. (2006). Design, delivery, and evaluation of training systems. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (3 ed., pp. 472-512). Hoboken, NJ: Wiley.

Schwid, H., Rooke, G., Carline, J., Steadman, R., Murray, W., Olympio, M., et al. (2002). Evaluation of anesthesia residents using mannequin-based simulation - A multiinstitutional study. *Anesthesiology, 97*(6), 1434-1444

Tannenbaum, S., & Yuki, G. (1992). Training and development in work organizations. *Annual Review of Psychology, 43*, 399-441.

Wickens, C., & Carswell, M. (1995). The proximity compatibility principle: its psychological foundations and relevance to display design. *Human Factors, 37*, 473-494.

Wickens, C., Helleberg, J., & Xu, X. (2002). Pilot maneuver choice and workload in free flight. *Human Factors, 44*(2), 171-188.