Measuring Pilot Knowledge in Training: The Pathfinder Network Scaling Technique

Leah J. Rowe
L-3 Communications
Mesa, AZ
leah.rowe@mesa.afmc.af.mil

Roger W. Schvaneveldt
Arizona State University
Mesa, AZ
schvan@asu.edu

Winston Bennett, Jr.
Air Force Research Laboratory
Mesa, AZ
winston.bennett@mesa.afmc.af.mil

ABSTRACT

Researchers have attempted to measure pilot knowledge and changes in knowledge, in both simulated and live-fly events. However, measurement in these training environments has been more successful in measuring overall flight performance outcomes rather than on underlying changes in knowledge. Research to assess changes in pilots' knowledge as a result of training is underway at the Air Force Research Laboratory (AFRL) in Mesa, Arizona, using the Pathfinder Network Scaling technique. The Pathfinder method uses individual judgments of the relationships between concepts/constructs in a domain as a basis to develop an empirically derived representation of knowledge about the concepts/constructs. These representations can be compared and changes in representation can be quantified to assess the impact of an intervention on knowledge. Previous research has demonstrated the value of Pathfinder for assessing the impact of both education and training interventions in domains such as computer programming. At AFRL, pilots, as part of a week-long 4-ship F-16 Distributed Mission Operations (DMO) training research program, participated in a Pathfinder study to asses F-16 pilot understanding of complex combat mission constructs/concepts critical to mission performance. The objective was to assess training effects that are more fundamental and process-orientated. This paper will report findings from a sample of 71 F-16 pilots who vary in experience level. Our results will be discussed both in terms of practical utility of the Pathfinder technique as a measurement methodology and in terms of knowledge measurement as a criterion for evaluating training.

ABOUT THE AUTHORS

Leah J. Rowe is a Training Research Specialist with L-3 Communications at the Air Force Research Laboratory, Warfighter Readiness Research Division in Mesa, AZ. She received her M.S. in Applied Psychology from Arizona State University in 2007.

Roger W. Schvaneveldt is the former Head and Professor of Applied Psychology at Arizona State University, Polytechnic Campus. He is one of the original developers of the Pathfinder network scaling technique. He holds a Ph.D. in Psychology from the University of Wisconsin.

Winston Bennett, Jr. is a Senior Research Psychologist and team leader for the training systems technology and performance assessment at the Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division, in Mesa AZ. He received his Ph.D. in Industrial/Organizational Psychology form Texas A&M University in 1995.

Measuring Pilot Knowledge in Training: The Pathfinder Network Scaling Technique

Leah J. Rowe
L-3 Communications
Mesa, AZ
leah.rowe@mesa.afmc.af.mil

Roger W. Schvaneveldt
Arizona State University
Mesa, AZ
schvan@asu.edu

Winston Bennett, Jr.
Air Force Research Laboratory
Mesa, AZ
winston.bennett@mesa.afmc.af.mil

INTRODUCTION

Simulated training events are beneficial to the military because they are less expensive and restrictive than live (non-simulated) training events. Establishing the validity of simulated events is an important criterion ensuring their continued use. Previous research in the Distributed Mission Operations (DMO) environment has shown that these operations improve F-16 flight performance across a variety of objective measures (Schreiber & Bennett, 2006a). If, in addition to improvements in performance measures, it can be demonstrated that knowledge measures display similar improvements, then the support for simulated training events is increased. The present research explores the role of knowledge structure in relation to performance during DMO. In the DMO environment, knowledge is measured using the Air Superiority Knowledge Assessment System (Gehr, Schreiber, Metz, & Bennett, 2005; Rowe, Gehr, Cooke, & Bennett, in press) and the Pathfinder Network Scaling technique. The present research explores the Pathfinder Network Scaling Technique in the Mesa, Air Force Research Laboratory (AFRL) DMO environment.

Pathfinder

Pathfinder is a knowledge elicitation technique developed in the 1980s (Schvaneveldt, Durso, & Dearholt, 1989). Since that time, Pathfinder has been applied to knowledge elicitation and representation in several domains. Some of the many applications include knowledge elicitation of military fighter pilots (Schreiber, DiSalvo, & Stock, 2006; Schvaneveldt, Tucker, Castillo, & Bennett, 2001), Air Battle Managers, Unmanned Aerial Vehicle teams (Shope, DeJoode, Cooke, & Pederson, 2004), anesthesiologists (Connor, Cooke, Weinger, & Slagle, 2004), and computer programmers (Cooke & Schvaneveldt, 1998).

Pathfinder extracts an underlying network from the judgments of individuals using mathematical graph theory. In mathematical graph theory, a graph consists of nodes and pairs of nodes (Harary, 1969). Each distinct pair of nodes is called a link. These links can be either directed or undirected. A set or group of nodes and links is then presented in the form of a graph with weights associated with the links. Taken as a whole, a collection of nodes and links can represent how an individual or a group views the relationships among concepts. An example of a network using general aviation terms is shown in Figure 1.

The links presented in the network are derived using individual judgments of the relatedness between all pairs of concepts. That is, each pair of concepts is numerically rated with respect to relatedness on a scale with “unrelated” on the lower end and “related” on the upper end.

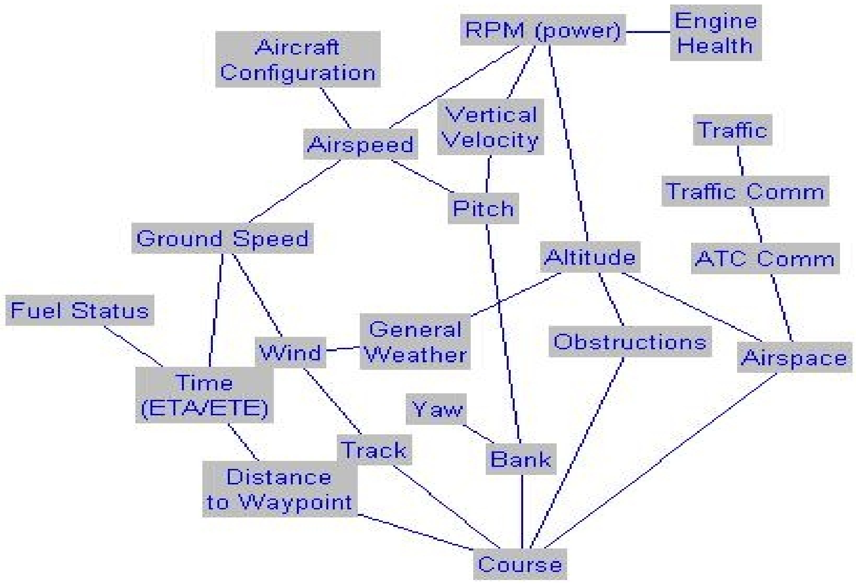


Figure 1. Pathfinder Network of General Aviation Terms

A substantial amount of research using the Pathfinder theory has taken place at AFRL. Previous research specifically focused on expert and novice ratings (Schvaneveldt, et al., 2001; Schreiber, et al., 2006). The analysis of pilot rating data includes measures of coherence and network similarity to experts. Coherence is a measure of the internal consistency of the ratings which often increases with growth in knowledge. The network similarity between individuals and experts provides a measure of the maturity of the knowledge structure of individual pilots. The present research focuses on the following research questions:

1. Will pilot coherence scores increase from the pre- to the post-assessment?

2. Will the participants' networks become more similar to the network of experts over time?

These questions were explored during Distrbuted Mission Operations (DMO) training research at AFRL.

Distributed Mission Operations (DMO)

DMO is a system of networked simulators that allow for multi-player training on combat exercises. DMO is different form stand-alone simulation systems, such as those used to train emergency procedures, in that it provides combat-like experiences involving real-time interaction with other entities, real (flight wingmen) and simulated (hostile entities).

The objective of DMO is to train higher-order skill development and teamwork coordination while executing significant portions of an entire mission (Colegrove & Alliger, 2002). Some DMO environments within the United States Air Forces include Shaw Air Force Base (AFB), Eglin AFB, Mountain Home AFB, and the AFRL Mesa Research Site in Mesa, AZ.

The environment for this study, AFRL Mesa Research Site, consists of four high fidelity F-16 simulators, a high fidelity Air Battle Manager simulator, a computer-generated threat system, and an instructor operator station. The F-16 simulators are labeled Viper 1 to 4. Vipers 1 and 3 are typically flight leads while Vipers 2 and 4 are wingmen. A well-equipped brief/debrief room is also available. Some features of the environment appear in Figures 2 and 3.

Figure 2. Overall view of Mesa AFRL DMO Training Research Environment

Figure 3. Interior view of a high fidelity F-16

simulator

METHODS

Participants

A total of 71 individuals, 15 teams of fully qualified F-16 United States Air Force, Air National Guard, or Air Force Reserve pilots participated in this study. Participants were between 24 and 44 years old, had between 3 and 23 years of experience, ranked between First Lieutenant (O-2) and Lieutenant Colonel (O-5), and had between 124 and 3600 F-16 flight hours. All participants volunteered. There was complete Pathfinder data for 61 of the 71 participants. Missing data was due to either incomplete data or equipment malfunctions.

An additional sample of experts was used as well. Six experts (from Schvaneveldt et al., 2001) completed the Pathfinder assessment using the same concepts as the participants did for the present study. These experts all possessed more than 1900 flight hours and all had high coherence scores (between .58 and .71).

Concepts Selection and Ratings

Pilots rated all pairs comprised from 21 different concepts thus producing a total of 210 relatedness judgments. The concepts were selected from advanced air-to-air combat maneuvering scenarios. To complete the ratings, the pilots used a numerical scale of one to nine where one was completely unrelated and nine was highly related. The concepts are listed in Table 1.

Table 1. Pathfinder air-to-air combat maneuvering concepts

Crank	Multiple Groups in Azimuth
AMRAMM	
Bandit/Hostile	Multiple Groups in Range
Beam Deploy	High Risk
BVR	PID
F-Pole	Pit Bull
Factor Bandit Range	Preserve Range
Grinder	Real World ROE
IRMD	Targeting/Sorting
Launch & Leave	Point Defense
MOR	Visual Mutual Support

Variables

In Pathfinder methodology, the q-parameter constrains the number of indirect proximities to generate the network. As q decreases the number of links added to the network increases. When analyzing individual proximity data it is recommended to use the q-parameters of n-1 (n is the number of nodes or rating items), and when averaging proximity data to use q=2 (Schvaneveldt, 1990). To compute the distance of paths the r-parameter is set to infinity in the case of ordinal data. For the present study the q-parameter was set to n-1 and r-parameter was set to infinity.

Pathfinder provides a coherence score, that is considered to be an index of internal consistency of the ratings, varying between 0 and 1. Pathfinder also produces network similarity scores for each participant that are based on the proportion of shared links between two networks. Two different pathfinder assessment scores were computed for this study to examine comparisons of individuals to a group of experts and to examine comparisons of individuals to an individual expert. The first score is the comparison of the individual networks for participants with the network derived from the average of the expert ratings. The second score is the comparison of the individual networks for participants with the network of the expert with the highest coherence score.

Performance

It recognized that the underlying purpose of simulated training is to increase the flight performance of the participants. Therefore, we also measured the flight performance of the participants. Each team's flight performance was measured using the Performance Evaluation Tracking System (PETS) (Schreiber & Bennett, 2006b). Performance was scored during two benchmark sessions, before and after DMO. The measures and scoring given in Table 2 were used to score each benchmark engagement at the team level.

Table 2. PETS Mission Performance Scoring Criteria

Event during benchmark	Performance Score Metric
Fratricide-Killed by blue air	-900
Mortality –Killed by red air	-300
Eliminate Striker- Kill striker prior to striker reaching base	+450 (900 possible per team of 4)
Elimination of Red Air	+150 (900 possible per team of 4)
Performance Score	Sum of points earned (1800 possible)

A strict protocol was employed during all benchmark scenarios to maintain a realistic combat environment and a consistent research environment. The benchmarks are point defense missions used to assess change in team performance from the beginning of the week to the end of the week. In total, there are seven different benchmark scenario pairs. Each scenario in a pair is the mirror image of the other scenario in the pair. Each team was randomly assigned three benchmark scenario pairs. Participants flew in the same cockpit position for all benchmark scenarios, on both Monday and Friday. Unknown to the participants, the mirror image of the three benchmarks flown on Monday were flown on Friday. The use of paired mirror-image scenarios ensures equivalent levels of difficulty and complexity during the Monday and Friday benchmark sessions. Figure 4 illustrates a benchmark and its mirror image. All of the benchmark scenarios that were utilized during this research have been established to have comparable levels of complexity (Denning, Bennett, & Crane, 2002).

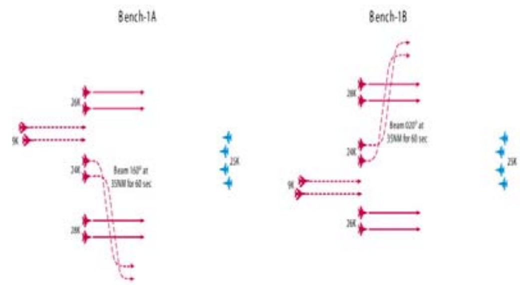


Figure 4. Example mirror image point defense benchmark scenarios used for the benchmark scenarios.

RESULTS

Pathfinder

A Pathfinder Network (PFNET) (r=infinity, q=n-1), was derived from each set of ratings for both before and after DMO assessments. Initially, the mean coherence for each Pathfinder participant assessment time (before and after) was analyzed. A paired t-test determined that coherence scores significantly increased from beginning (M = 0.448) to end (M = 0.497) of the DMO training (t(60)=2.01, *p=.02*), see Figure 5.

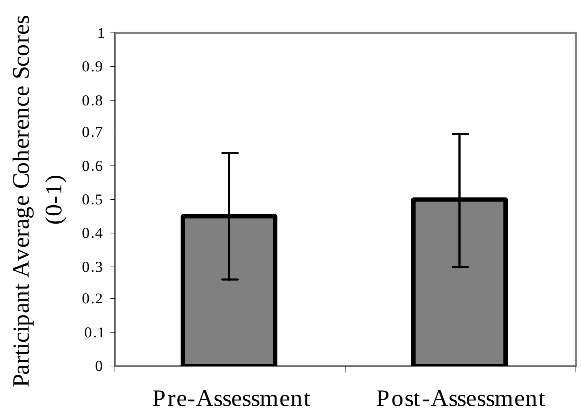


Figure 5. Pathfinder Pre- and Post- DMO Assessments Coherence scores

Furthermore, the correlation with the expert with the highest coherence score significantly increased from before DMO (Mean correlation = .325) to after (Mean correlation = .347) (t(60)=1.84, *p=.03*) (see Figure 6), but no significant difference existed when the correlation was calculated using the average of experts in the paired t-test (t(60)=1.30, *p=.09)*.

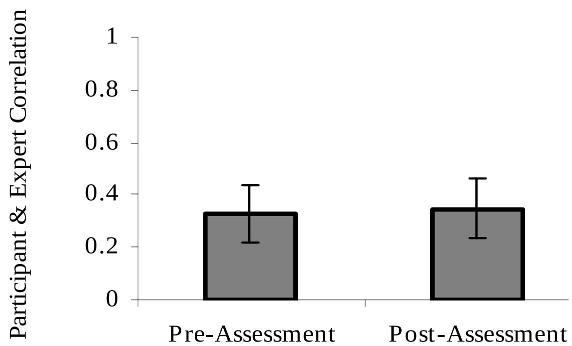


Figure 6. Correlation between one expert and participants' PFNET for the initial and final Pathfinder assessments

The remainder of the Pathfinder analyses compared the participant's ratings to expert ratings. It was determined that the participants had more of their weighted links in common with experts at the end of the week (47.61%) than at the beginning of the week (33.33%) as shown in Figure 7.

Flight Performance

A paired t-test determined the average flight performance significantly increased from before to after the training with an initial performance mean score of 1,250 (SD = 346.41) and final mean score of 1,578.12 (SD = 324.02) (t (14) =3.68, *p*<.05), as shown in Figure 8.

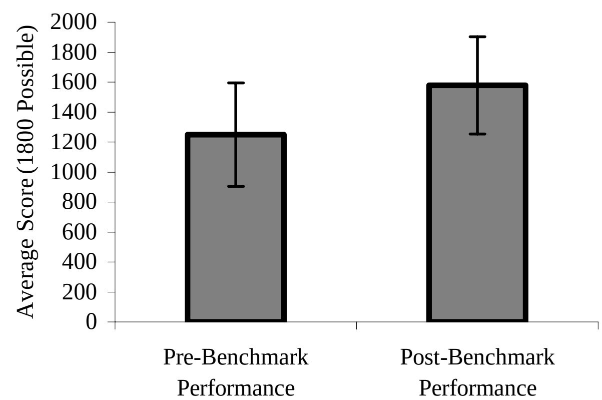
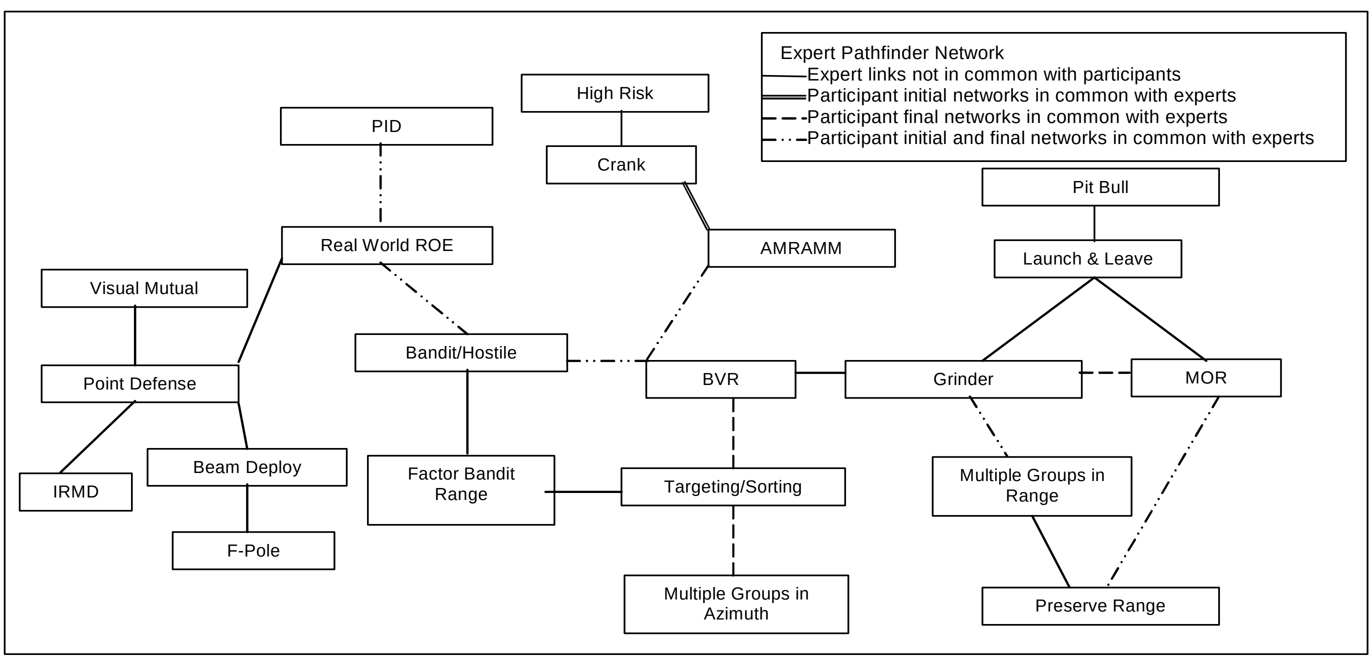


Figure 8. Before and after DMO benchmark flight performance scores

Figure 7. Participants networks in common with expert networks before and after DMO



DISCUSSION AND CONCLUSION

Consistent with flight performance scores, training led to a significant increase in the similarity between participant networks and the network from the expert with the highest rating coherence. While individual networks became more similar to experts using the network derived from the average of expert ratings, it was not significant. Perhaps comparisons between individual networks leads to a more sensitive index because such comparisons do not average out important factors for evaluating knowledge change. This finding deserves more study.

DMO training is heavily dependent on a team of pilots. Whereas the present flight performance metrics aim at the team as a unit, the knowledge assessment tools only consider the individual. To address the relationship between DMO flight performance and knowledge acquisition, knowledge should also be measured at the team level, along with other team measures like cohesion.

In future knowledge acquisition studies it would be useful to use a team Pathfinder rating system rather than to aggregate or average individual scores to get a team score. This would allow the team of participants to communicate regarding their ratings prior to inputting a rating, encouraging them to share information among the team. In a DMO type of environment this rating system would enhance the team as a unit allowing each individual to have a better understanding of each other's strengths and weaknesses in their given roles and with their levels of expertise.

ACKNOWLEDGEMENTS

This research was performed at the Air Force Research Laboratory, Warfighter Readiness Research Division in Mesa, AZ, under Air Force contract 8650-05-D-6502, Principle Investigator Dr. Winston Bennett, Jr.

REFERENCES

Colegrove, C. M., & Alliger, G. M. (2002). Mission essential competencies: Defining combat readiness in a novel way. In *Proceedings of the NATO Research & Technology Organization, Studies, Analysis, and Simulation Panel, Conference on Mission Training via Distributed Simulation* (SAS-38), Brussels, Belgium.

Connor, O., Cooke, N. J., Weinger, M. B., & Slagle, J. (2004). Using psychological scaling techniques to assess clinical expertise in anesthesiology. In *Proceedings of the Human Factors and Ergonomics Societies 48th Annual Meeting*, 1746, 1750. New Orleans, LA: Human Factors and Ergonomics Society.

Cooke, N. J., & Schvaneveldt, R. W. (1988). Effects of computer programming experience on network representations of abstract programming concepts. *International Journal of Man-Machine Studies*, 29, 407-427.

Denning, T., Bennett, W. Jr., & Crane, P. M. (2002). Mission Complexity Scoring in Distributed Mission Training. In 2002 *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*. Orlando, FL: National Security Industrial Association.

Gehr, S. E., Schreiber, B. T., Metz, C., & Bennett, W. Jr. (2005). Air Superiority Knowledge Assessment System. In 2005 *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*. Orlando, FL: National Security Industrial Association.

Harary, F. (1969). *Graph theory*. Reading, MA: Addison-Wesley.

Rowe, L. J., Gehr, S. E., Cooke, N. J., & Bennett, W. Jr. (in press). Assessing Distributed Mission Operations Using the Air Superiority Knowledge Assessment System. In 2007 *Proceedings of the Human Factors and Ergonomics Annual Meeting*. Baltimore, MD.

Schreiber, B. T., & Bennett, W. Jr. (2006a). *Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study. Volume I: Summary Rep*ort (AFRL-HE-AZ-TR-2006-0015-Vol I). Mesa AZ: Air Force Research Laboratory, Warfighter Training Research Division.

Schreiber, B. T., & Bennett, W. Jr. (2006b). Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study. *Volume II: Metric Development and Objectively Quantifying the Degree of Learning* (AFRL-HE-AZ-TR-2006-0015-Vol II). Mesa AZ: Air Force Research Laboratory, Warfighter Training Research Division.

Schreiber, B., DiSalvo, P., Stock, W. A., & Bennett, W. Jr. (2006). Distributed Mission Operations Within-Simulator Training Effectiveness Study. *Volume V: Using the Pathfinder Methodology to Assess Pilot Knowledge Structure Changes* (AFRL-HE-AZ-TR-2006-0015-Vol III). Mesa AZ: Air Force Research Laboratory, Warfighter Training Research Division.

Schvaneveldt, R. W. (Ed.) (1990). *Pathfinder associative networks: Studies in knowledge organization.* Norwood, NJ: Ablex.

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. In G. Bower (Ed.) *The psychology of learning and motivation: Advances in research and theory*, Vol. 24 (pp. 249-284). New York: Academic Press.

Schvaneveldt, R. W., Tucker, R., Castillo, A., & Bennett, W., Jr. (2001). Knowledge acquisition in distributed mission training. In 2001 *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC).* Orlando, FL: National Security Industrial Association.

Shope, S. M., DeJoode, J. A., Cooke, N. J., & Pederson, H. (2004). Using Pathfinder to generate communication networks in a cognitive task analysis. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, 678-682. New Orleans, LA: Human Factors and Ergonomics Society.