2009

Assessing Novel Adaptive Displays Impact on Pilot Performance

Sean L. Guarino

Karen Harper

Emilie Roth

Dahai Liu

Mark Bartosiewicz

See next page for additional authors

ASSESSING NOVEL ADAPTIVE DISPLAYS IMPACT ON PILOT PERFORMANCE

Sean L. Guarino, Karen Harper
Charles River Analytics Inc.
Cambridge, MA 02738

Dahai Liu, Mark Bartosiewicz
Embry-Riddle Aeronautical University
Daytona Beach, FL 32114

Emilie Roth
Roth Cognitive Engineering
Brookline, MA 02445

Dennis Vincenzi
Naval Air Warfare Center Training Systems Division
Orlando, FL 32826-3275

Combat aircraft advances have led to a dramatic increase in the operational tempo facing the Navy pilot, increasing the likelihood for Situation Awareness (SA) failures, biased decision-making and information overload. We designed a system for constructing intelligent adaptive displays to address these issues and, within it, designed and evaluated two interfaces targeted at tactical SA challenges: (1) a *Weapons Employment Zone (WEZ) Display* designed to support awareness of combat geometry; and (2) an *Adaptive Boundary Display (ABD)* designed to warn pilots of impending border infractions that would compromise Rules of Engagement (ROEs). We tested the ability of these displays to improve SA, reduce workload, and improve mission performance in a population of licensed civilian pilots. The WEZ Display significantly improved performance and SA and reduced workload, while the ABD made no significant improvements. We recommend using the WEZ Display to assist novice pilots in understanding and tracking real-time combat geometry.

Introduction

Advances in aircraft performance and operational capabilities have led to a dramatic increase in the operational tempo facing the air combat aviator, reducing the available time to process larger sets of tactically-relevant information and make effective operational decisions based on that information. The technological and information advances of Network Centric Warfare (NCW) have resulted in an explosion in the quantity and complexity of information available to aviators (Shinseki & Caldera, 1999) who must track, monitor, and process information arriving from many disparate information sources and integrate that information into their cognitive decision-making processes (Endsley & Bolstand, 1992). To counter these increasingly complex operational environments, we must develop advanced Human/System Interface (HSI) capabilities that will make optimal use of human operators' cognitive resources.

Modern HSIs do little to address the complexities associated with aircraft operations. While they present the basic information needed for decision-making, they neglect to do so in a manner supporting a proper understanding of the ongoing situation. Most systems do little (if anything) to restrict information flow, regardless of the potential for information overload. Furthermore, most systems neglect to evaluate operator activities, and do nothing to address potential errors caused by decision-making biases. Human decision-making errors that occur when interacting with non-adaptive HSIs can be grouped into three categories: 1) **Situation Assessment Failures**, in which operators fail to detect or correctly interpret the information presented to them, and therefore miss key cues and events central to the formation of an accurate mental model of the situation; 2) **Workload Failures**, in which information overload can lead to information loss and oversight; and 3) **Decision Bias Failures**, in which natural human biases (e.g., a framing of the situation at hand, salience, cues, or heuristics intended to enable rapid decision making) can lead the operator to misinterpreting the tactical situation. To counter these increasingly complex issues, we have developed a middleware system, the *Modular Adaptive Interface Suite* (MAIS), that supports the development of intelligent systems to drive adaptive HSIs, supporting adaptations that optimally use human operators' cognitive resources and promote the formation of accurate mental models of the situation.

In this effort, we used Cognitive Task Analysis (CTA) methods (Mahoney et al., 2008; Bisantz & Roth, 2008; Pfautz & Roth, 2006; Schraagen, Chipman, & Shalin, 2000) including structured interviews, to identify a number of key SA issues arising in air combat aviation. This CTA involved extensive interviews with one former fighter pilot, as well as observations of fighter pilots during training and interviews with current instructors. Several issues were identified, two of which were the focus of the current research: 1) correctly monitoring Rules of Engagement (ROEs) and the location of key geopolitical boundaries to avoid geopolitical incidents during operations near borders; and 2) understanding combat geometry and monitoring the location of the enemy Weapons

Employment Zone (WEZ) and the threat posed to ownship, wingmen and escorted strikers. We designed, developed, and performed a study to analyze display adaptations designed to address these two issues.

First, we developed an *Adaptive Boundary Display* (ABD), designed to address issues aviators have in monitoring geopolitical boundaries while engaging hostile foreign aircraft that attempt to draw them across borders. If provided the opportunity in such situations, the hostile entities will fire missiles. Otherwise, they will attempt to draw military aviators into illegally entering the foreign territory, thus creating a geopolitical incident. Despite ROEs that clearly state the restriction of not crossing the border, these threats can cause the military aviator to lose track of where the border is and inadvertently cross it. To address this issue, we developed an ABD within the Combat Situation Display of the aircraft which dynamically changes the color of the boundary based on the analyzed likelihood that the aviator is going to pass the boundary. The boundary, initially shown in gray, changes to a bright purple and eventually turns red as the need for a reaction from the aviator increases.

Next, we developed a *WEZ Display*, designed to assist aviators in forming an accurate mental model of the combat geometry and understanding the threat posed by enemies. The WEZ defines the missile envelope in front of the aircraft, splitting it into areas in which an enemy is likely to be eliminated and in which an enemy is likely to escape. Realistically, the WEZ is best represented as a bubble in front of the aircraft, as illustrated in Figure 1A. To simplify this representation for aviators, our WEZ display shows a two-dimensional cone representation of the area in front of aircraft (see Figure 1B). The WEZ is characterized in four regions: a *No-Shot* region directly in front of the aircraft, in which the target cannot be safely fired upon; a *No-Escape* region beyond that, in which the target cannot physically escape the missile; a *High Probability of Kill (PK)* region in which a target is unlikely to escape the missile; and a *Low PK* region in which the target's chance to escape is higher, but still not certain. The goal of the military aviator is to achieve a High PK or No Escape shot before providing the opponent with a First Launch Opportunity (FLO) (see Figure 1C). To address this issue, we developed a *WEZ Display* within the Combat Situation Display of the aircraft, showing a constant two-dimensional representation of the known ownship WEZ in front ownship, and a worst-case representation of the enemy WEZ in front of each enemy aircraft icon.

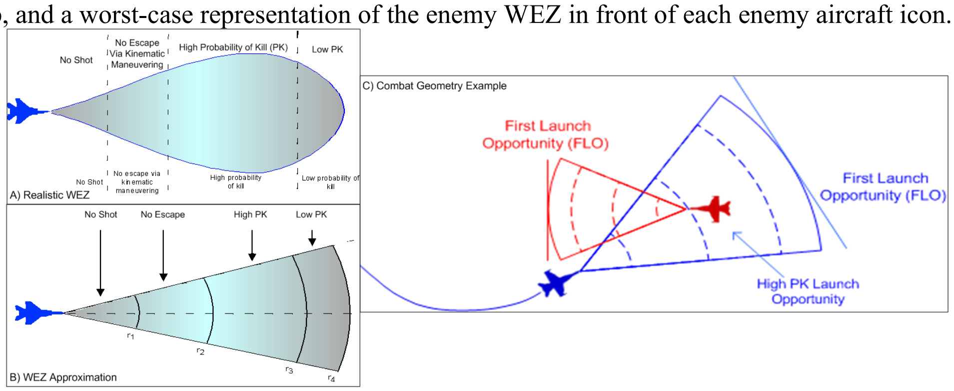

Figure 1: Weapons Employment Zone: A) Realistic WEZ; B) WEZ Approximation; C) Combat Geometry Example

We designed a study to test these adaptive displays as described below. For this study, we made the following hypotheses: 1a) the WEZ Display will improve overall pilot performance by decreasing the enemy FLOs and increasing average shot PK; 1b) the WEZ Display will increase SA; 1c) the WEZ Display will reduce workload; 2a) the ABD will improve overall pilot performance by decreasing instances of border crossings; 2b) the ABD will increase SA; 2c) the ABD will reduce workload; and 3) combining the ABD and WEZ Display will create an improved interaction effect over either display alone.

Methods

Participants: Sixty U. S. citizen participants (54 male, 6 female, average age 22) were recruited from the university community to participate in the study. Participants were all experienced civilian student pilots (59 have private pilot licenses), with an average of 67.3 hours of simulation training, 177.7 VFR hours, 47.3 IFR hours and 57.7 hours of Technological Advanced Aircraft (TAA, i.e., glass cockpit) experience. Most pilots have no air combat experience, although some have related gaming experiences (with an average 5 hours in those types of games).

Experimental Apparatus: In this experiment, we implemented our adaptive displays in a version of OpenEaagles (the Open Extensible Architecture for Analysis and Generation of Linked Simulations, http://openeaagles.org/). Our version of OpenEaagles provides a medium-fidelity air combat simulation environment, including a representation of the Heads Down Display (HDD) of an F-16 with basic instruments and multifunctional displays. We implemented our WEZ Display and Adaptive Border Display as optional components within the OpenEaagles

Tactical Situation Display (TSD). The TSD shows the aviator's ownship at the center of an overhead view and also shows any detected aircraft that are within range of that view. Figure 2A shows a representation of the TSD, and Figure 2B, C, and D shows representations of the Adaptive Border Display, WEZ Display, and combined displays, respectively. Additionally, we enhanced OpenEaagles to record flight path data and a number of key events, such as entering the enemy WEZ, crossing the boundary, or firing a missile at the enemy. In our experimental version of OpenEaagles, each trial immediately ended in one of two cases: 1) when the pilot fired a missile with a PK greater than zero at each opponent, or 2) when the pilot or the escorted striker entered an enemy WEZ.

Figure 2: Tactical Situation Display: A) Basic TSDs; B) TSD w/ ABD; C) TSD w/ WEZ Display; D) TSD w/ ABD and WEZ Display

We implemented an experimental aid component that managed the experiment. Between scenarios, this dialog reported the trial performance to the participant and reminded the participant to fill out between-trial SA and workload measures.

Experimental Tasks and Procedures: Participants first read and signed an informed consent form and filled out a demographic form. Upon completion, they were given a training session on the simulation environment, controls, and concepts of combat geometry and ROEs. Each participant flew an uncontrolled flight scenario to become familiarized with the controls of the simulation environment; this scenario provided no enhanced displays and no border conditions, although it did provide combat situations to familiarize the participant with combat controls. Participants then ran a practice training trial to become familiarized with the format of an experimental trial. This was a three to five minute combat trial that did not count in their final score, in which the participants flew a combat scenario, and performed SA and workload tasks after the scenario. In this trial, participants were also introduced to any additional display features that were available in their condition.

After a break, each participant went through twelve experiment trials. For each trial, the participant flew a fighter aircraft on striker escort missions in a fictional country, called Targetzistan, where his/her job was to protect the striker. Participants were expected to engage in air-to-air combat maneuvers against enemy aircraft coming from Aggressistan. The twelve trials varied in complexity level, following four templates illustrated in Figure 3. Each trial lasted approximately three to seven minutes, and was followed by a SA reconstruction task and a partial NASA-TLX task. After the twelve trials, participants were asked to fill out an overall NASA-TLX survey and an SA questionnaire. It took approximately three hours for one participant to complete the tasks.

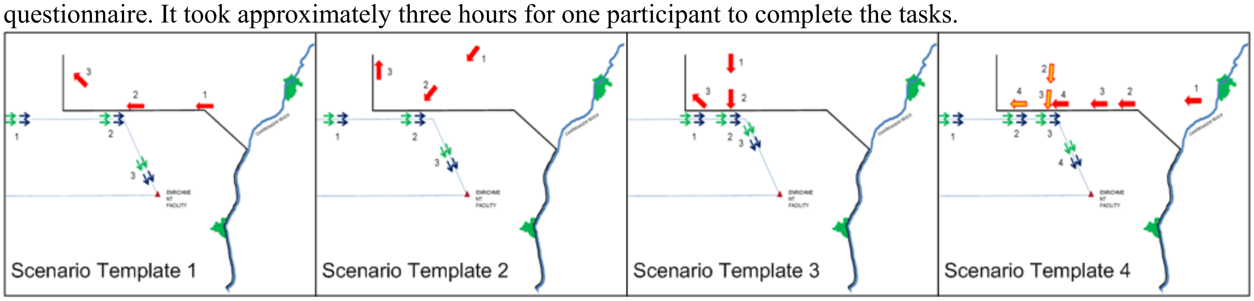

Figure 3: Experimental Scenario Templates

The experiment utilized a 2x2 between-subject design, in which experimental factors were (1) use of the Adaptive Border Display and (2) use of the WEZ Display. This 2x2 design resulted in four display conditions (Adaptive Border Display only, WEZ Display only, both displays, and a control case of no extra displays. Participants were randomly assigned to one of four display conditions (15 participants in each condition).

Data Collection: Three dependent variables were collected, including objective performance, SA, and workload. For each trial, our software automatically measured objective performance based on logged events. This objective performance metric evaluated the participant's success in:

1. **Avoiding breaks in rules of engagement:** -5 points for each instance of breaking the rules of engagement (e.g., each entry into foreign territory).
2. **Minimizing necessary breaks in rules of engagement:** -1 points per every five seconds in foreign territory.
3. **Denying enemy first launch opportunities:** +10 points for denying the enemy first launch opportunity (e.g., avoiding the enemy WEZ and firing on the enemy before the enemy WEZ encompasses the escorted striker).
4. **Achieving an effective launch on enemies:** +0 to +10 points for a launch taken at the enemy, calculated based on the probability of kill (PK) of the shot (0.0 for a 0% PK, +10 for a 100% PK); when there are multiple targets in a trial, this is the mean of the best PK shot taken on each enemy

Note that during training, participants were explicitly informed of the details of the performance metrics, and informed that awards would be based on maximizing performance. The overall performance score was computed by using the four metrics listed above; meanwhile, individual performance metrics were also captured, including Enemy First Strike Opportunities (measuring if the participant or the escorted striker enter the enemy WEZ before the participant makes an effective shot); Shot Probability of Kills (PK) (measuring the probability of shooting down the enemy aircraft) and Boundary Infractions (measuring number of times crossed the border boundary).

SA was measured using an SA reconstruction task designed to assess the level of situation retention after the tasks were completed. In this task, after each trial, participants drew a re-enactment of the scenario, attempting to recall what occurred during the trial. By comparing that data to what really occurred based on our log files, we were able to answer the following five questions:

- Did the participant correctly identify and interpret ownship border crossings?
- Did the participant correctly identify and interpret enemy aircraft border crossings?
- Did the participant correctly identify and interpret enemy first strike opportunities on ownship?
- Did the participant correctly identify and interpret enemy first strike opportunities on escorted aircraft?
- Did the participant correctly identify and interpret own WEZ position?

Based on the answers to these questions, we rated participants from 0-5 to produce an SA score for each trial. Additionally, at the close of the experiment, we assessed overall self-assessment of SA using an SA questionnaire. This questionnaire included fourteen questions using a 0-10 Likert scale to measure participants' perceptions of their own SA. There were 14 questions total in this questionnaire, six were relative to the SA of the geopolitical boundary and eight were related to SA of ownship and enemy WEZ. Score was averaged across the fourteen questions to provide a composite measurement for SA.

Finally, workload was assessed through the NASA-TLX workload rating scale (Hart & Staveland, 1988). The NASA TLX is a subjective workload measure that captures participant's ratings of the mental demand, physical demand, temporal demand, own performance, effort, and frustration level while performing a task. The standard NASA TLX survey is two pages: in the first, participants rate the workload experienced by each of the six specified scales; in the second, they weigh which of the factors they consider more important to measuring workload by performing a pairwise comparison. To shorten the task between trials, we only had participants rate the six scales at the end of each trial. Then, once, at the end of the experiment, they filled out the full NASA TLX for the full experiment.

Results

A between subject Analysis of Variance (ANOVA) was conducted on the three dependent measure. Table 1 summarizes the ANOVA results on all the dependent measures. Based on these results, it is clear that the WEZ display had a significant effect on overall performance score (p=0.008, power of 0.769) and on reducing enemy First Strike Opportunities (p=0.040, 0.541). Figure 4A and 4B illustrate these performance differences for the WEZ Display. Unfortunately, the ABD and interaction between the WEZ and ABD were not significant. Similarly, there was no significant effect from either display on PK of shots taken or on boundary infractions (although, the PK of shots did show a non-significant trend towards improvement when using the WEZ Display (p=0.064, power = 0.459)).

There were two parts of our SA analysis. For individual trials, we analyzed the SA reconstruction task, counting the wrong answers in each case, and finding no significant effects. Additionally, after the experiment, we analyzed an SA questionnaire to access participants' subjective impression of the level of SA provided by their condition. Here, the WEZ had a significant effect on overall SA impression (p=0.013, power = 0.708), and the APD remained non-significant. Figure 4C illustrates these subjective SA differences.

Finally, NASA-TLX measurements were taken to gather subjective impressions of workload during each trial, and in the overall experimental condition. As with SA, the WEZ Display significantly reduced participants' perception of overall workload (p=0.013, power = 0.715), but not of workload during individual trials (although it approached significance, with p=0.051).. Again, the APD remained non-significant. Figure 4D illustrates these subjective SA differences.

Table 1: *ANOVA Analysis Results*

Dependent Measures	**p_{WEZ}**	**$Power_{WEZ}$**	**P_{ABD}**
Overall Performance	0.008*	0.769	0.956
First Strike Opportunities	0.040*	0.541	0.487
Shot PK	0.064	0.459	0.385
Boundary Infractions	0.587	0.084	0.416
Trial SA (Reconstruction Task)	0.238	0.216	0.823
Overall SA (Questionnaire)	0.013*	0.708	0.195
Trial TLX	0.051	0.501	0.644
Overall TLX	0.013*	0.715	0.817

Note. An asterix (*) indicates significant effect.

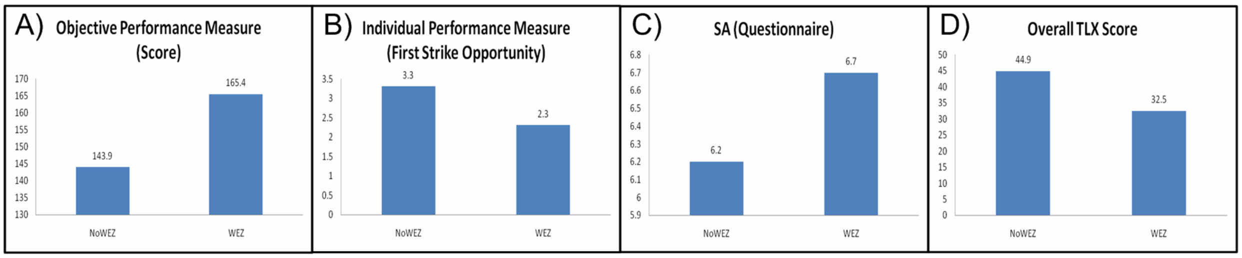

Figure 4: A) Effect of WEZ Display on overall performance, significant at p=0.008; B) Effect of WEZ Display on FSO, significant at p=0.040; C) Effect of WEZ Display on overall SA, significant at p=0.013; D) Effect of WEZ Display on Overall Workload, significant at p=0.013

Discussions and Conclusions

In this study, we investigated two adaptive concepts: the ABD and the WEZ Display. Our first group of hypotheses was that the WEZ Display would enhance performance and SA and reduce workload. Each of these hypotheses was validated by our results. Results showed that the WEZ had a significant effect across most of the dependent measures, including performance measures, overall SA, and overall workload measures. In improving performance, the WEZ Display in particular helped pilots to prevent enemy FLOs. Additionally, it appears that the WEZ Display created a trend of improvement in shot PK (p=0.064). Overall, this evidence indicates that the WEZ Display successfully provided real-time information on the area threatened by enemy aircraft, allowing the pilot to both immediately recognize when his escorted striker was threatened, and maneuver to achieve a high-PK shot without entering the enemy WEZ. With this information aid, aviators can quickly form a more accurate mental model of the combat geometry, enabling them to make a more prompt and accurate response to the situation. Additionally, based on our workload results, the WEZ Display appears to reduce the mental effort required to understand the combat situation and make decisions within the posed combat scenarios.

Our second group of hypotheses was that the ABD would enhance performance and SA and reduce workload. Unfortunately, the ABD had no significant effect on any of these dependent variables. In hindsight, this data is reflective of information we learned in the CTA process, where our SMEs suggested that the issue with monitoring the geopolitical boundary during an engagement occurred primarily because enemies would create a complex "cat and mouse" type of engagement, attempting to draw pilots into geopolitical incidents. In these cases, enemy aircraft would remain just outside of missile range until the pilot was not attending to them and then quickly move in for a strike. In light of this information, we believe that the findings of this study are non-significant because we failed to create this complex type of engagement. In particular, geopolitical incidents only occurred in 30 times across the 720 trials run on all participants in this study. In a follow-up study, we will further investigate the ABD, using more complex scenarios where enemy aircraft attempt to induce the pilots to create an incident by crossing the border, and where enemy WEZ sizes are more threatening.

Our final hypotheses were that the WEZ Display and ABD would combine to enhance both performance and SA and reduce workload over either enhancement alone. Unsurprisingly, the ABD also provided no improvement over the WEZ Display condition. Again, we believe this is because we were not creating scenarios that require the ABD.

Limitations and Future Research

While our results successfully illustrated that pilots can benefit from our WEZ Display, there are a number of limitations on our results. First, as mentioned above, our experiment did not provide the situations required to effectively test the ABD. In a future study, we will test the ABD with more complex scenarios designed to create situations where the ABD would be useful.

Next, our population consisted of a group of civilian pilots, not air combat pilots. By using a population of student pilots at Embry Riddle Aeronautical University, we had experienced pilots, capable of effectively flying scenarios in OpenEaagles with minimum training time. However, while we provided our pilots with initial training in air combat tactics, even Navy novices have more education in recognizing the details of actual combat situations. Clearly, our experiment would be more convincing if we had a military pilot population, and we recommend future research to investigate such a population. Even a small population of retired military pilots showing similar results for WEZ Display would be useful.

A third limitation of our research was the realism of the experimental system. While OpenEaagles provides a solid medium-fidelity air-to-air combat simulation environment, it clearly does not capture all of the complexity of a real aircraft. To begin with, in our experiment we only included a Heads Down Display (HDD) to simplify the task for our non-combat pilots; in a real aircraft, pilots would of course also have an out the window display and a Heads Up Display (HUD), each of which could distract them from our adaptive display. Furthermore, our TSD had no underlying map. Such displays in real aircraft often have a map in the background. This is a significant limitation because it is unclear what effect the map would have on observing both static and adaptive boundary lines. Clearly, performing this experiment in a high-fidelity simulation environment would produce more reliable results.

Nevertheless, based on our results, the WEZ Display appears to be a useful feature for increasing pilot performance in complex air combat situations, helping pilots to understand complex combat geometry problems. We recommend future research investigating the application of these adaptations in more realistic simulation environments and, ultimately, in existing and future military aircraft, and with a population of Navy aviators. We also recommend the development of further adaptations to address other key SA and workload issues plaguing Navy aviators, such as recognition of potential aircraft energy issues, flow issues, and combat timeline issues. Finally, we recommend the application of our intelligent adaptive display technology to development in associated high-tempo domains, such as rotary aircraft piloting or unmanned air vehicle (UAV) piloting.

Acknowledgements

This work was performed under US Navy contract number N61339-07-C-0025 with the Navy Air Warfare Center Training Systems Division (NAWCTSD). The authors thank Dr. Dennis Vincenzi and Mr. Richard Plumlee for their support and direction on this project, and their aid in finding SMEs for our CTA process. Additionally, the authors thank LCDR Michael McLeod (Ret.) for serving as our central SME in this process. Finally, we would like to thank Mr. Randy Levine from ASC/SIMAF for providing support in our use of OpenEaagles for this effort.

References

Bisantz, A. & Roth, E. (2008). Analysis of Cognitive Work. *Review of Human Factors and Ergonomics,* 3.

Endsley, M. R. & Bolstand, C. A. (1992). Human Capabilities and Limitations in Situation Awareness. In *Proceedings of AGARD: Combat Automation for Airborne Weapons Systems: Man/Machine Interface Trends and Technologies*. Neuilly Sur Siene, France.

Hart, S. G. & Staveland, L. E. (1988). The Development of the NASA Task Load Index (TLX): Results of Empirical and Theoretical Research. In P. A. H. Meshkati (Ed.), *Human Mental Workload* (pp. 139-183). North Holland, Amsterdam.

Mahoney, S., Pfautz, J., Roth, E., Fichtl, T., Powell, G., Guarino, S. et al. (2008). Cognitive Systems Engineering for Evolvable Human-in-the-Loop Data Fusion. In *Proceedings of 52nd Annual Meeting of the Human Factors and Ergonomics Society*.

Pfautz, J. & Roth, E. (2006). Using Cognitive Engineering for System Design and Evaluation: A Visualization Aid for Stability and Support Operations. *International Journal of Industrial Engineering,* 36(5), 389-407.

Schraagen, J. M. C., Chipman, S. F., & Shalin, V. L. (2000). Introduction to Cognitive Task Analysis. In J. M. C. Schraagen, S. F. Chipman, & V. L. Shalin (Eds.), *Cognitive Task Analysis* (pp. 3-24). Mahwah, NJ: Lawrence Erlbaum Associates.

Shinseki, G. E. K. & Caldera, L. (1999). *The Army Vision Statement.* US Army.