The Virtuous Circle and Contextualized Knowledge Elicitation: Application of a New Paradigm for Job Analysis

Sterling Wiggins[1], Jennifer Pagan[2], Michael J. Keeney [1], Amy Bolton[3], Webb Stacy[1], Jeffrey M. Beaubien[1]

[1]Aptima, Inc.
Woburn, MA 01810
swiggins@aptima.com , mkeeney@aptima.com wstacy@aptima.com jbeaubien@aptima.com

[2]Naval Air Warfare Center, Training System Division
Orlando, FL 32826
jennifer.pagan@navy.mil

[3]Office of Naval Research
Arlington, VA 22203
amy.bolton@navy.mil

ABSTRACT

Landing on the deck of an aircraft carrier is one of the most difficult and dangerous tasks that a Naval aviator must perform. *Maritime Augmented Guidance with Integrated Controls for Carrier Approach and Recovery Precision Enabling Technologies* (MAGIC CARPET) is a technology intended to make carrier landing performance better and safer, and to reduce pilot workload. The workshop described in this report included a set of contextualized knowledge elicitation sessions that provided preliminary multi-faceted evidence that MAGIC CARPET leads to improved performance during landings and to lower participant workload. Contextualized knowledge elicitation collects a combination of system-generated data and self-reports together in real-time. The combination of system-generated data and self-reports can provide cues to guide investigations about perceptual-cognitive skills required for successful task performance in high-consequence environments. Self-reports can highlight where to look within voluminous system data to gain insights about operator or system performance. Similarly, system data can indicate where verbal follow-ups can provide additional context about the objective data that were collected. The workshop was held at the Manned Flight Simulation operation at Naval Air Station Patuxent River. Six Naval Aviator participants, representing various F/A-18 experience levels, plus two non-pilot engineers used the simulator to conduct carrier landings using both conventional and MAGIC CARPET technology. Participant and Landing Signal Officer (LSO) reports as well as physiological and simulator-based measures showed a strong, noticeable, positive effect of MAGIC CARPET on landing performance. Participants reported dramatically reduced workload and LSOs reported that MAGIC CARPET made less-experienced participants perform as well as experienced pilots. These subjective participant observations were borne out by objective measures. Finally, the workshop provided rich information about a future more formal experiment to assess the impact of MAGIC CARPET on pilots and on the pilot training pipeline.

ABOUT THE AUTHORS

Sterling Wiggins is a Principal Cognitive Scientist in the Performance Assessment and Augmentation division at Aptima, Inc. He leads several projects creating technology and training to support humans working in high risk, safety-critical environments. His research interests include Live, Virtual, and Constructive training, human-automation interaction, and adaptive aiding. He holds an M.A in Education with a focus on learning, design, and technology from Stanford University, and a B.S. in Psychology from Ohio State University.

Jennifer Pagan is a Research Psychologist at the Naval Air Warfare Center Training Systems Division (NAWCTSD) Orlando, FL. She currently manages a research and development effort devoted to investigating automated performance and physiological measures. Additionally, she is the Principal Investigator on multiple efforts investigating selection concerns for unmanned systems. She currently supports several research and development efforts focused on instructional system capabilities and Live, Virtual, and Constructive training. Her research interests include performance assessment and measurement, personnel selection, test validation, and training effectiveness evaluation. She holds an M.S. degree in Industrial/Organizational (I/O) Psychology from the University of Central Florida.

Dr. Michael J. Keeney has extended the application of methodology traditionally used for adaptive testing to use in adaptive training tools. During military service, Dr. Keeney re-engineered two blocks of technical instruction, reducing student attrition while enhancing the value of the instruction, and his improvements to selection tests and a second technical course resulted in selection for an Air Force-level award. Dr. Keeney holds a Ph.D. in Industrial/Organizational Psychology from the University of Akron, an M.A. in Psychology from The University of Akron, a B.A. in Psychology from The University of Maryland, and an A.A. in Technical Management from the Community College of the Air Force. He is a member of American Psychological Association and the Society for Industrial and Organizational Psychology.

Dr. Amy Bolton is a Program Officer at the Office of Naval Research (ONR). She manages several programs within the Capable Manpower Future Naval Capability. Capable Manpower is a multi-million dollar per year initiative that addresses manpower, personnel, training, and human system design Science and Technology challenges for the Navy and Marine Corps. Dr. Bolton's research interests include adaptive training, human behavior modeling, human system design, and Live, Virtual, and Constructive training. She holds a Ph.D. in Applied Experimental and Human Factors Psychology from the University of Central Florida.

Dr. Webb Stacy is a Corporate Fellow at Aptima, Inc., where he is responsible for enhancing Aptima's technology portfolio. Dr. Stacy has an interest in using modern Cognitive Science to improve experiential training. His recent work includes investigating the relationship of simulator fidelity to training effectiveness and developing an approach to optimizing the training value of experiential scenarios. Dr. Stacy holds a Ph.D. in Cognitive Science from SUNY/Buffalo, a B.A. in Psychology from the University of Michigan, and is Program Chair-elect for the Society for Behavior Representation in Modeling and Simulation.

Jeffrey M. Beaubien is a Principal Scientist in the Advanced Cognitive Training Systems division at Aptima, Inc. For the past 15 years, his work has focused largely on training and assessing leadership, teamwork, and decision-making skills. His work has been sponsored by the U.S. Navy, the U.S. Army, the U.S. Air Force, the Federal Aviation Administration, and the Telemedicine and Advanced Research Technologies Center, among others. Dr. Beaubien holds a Ph.D. in Industrial and Organizational Psychology from George Mason University, an M.A. in Industrial and Organizational Psychology from the University of New Haven, and a B.A. in Psychology from the University of Rhode Island.

The Virtuous Circle and Contextualized Knowledge Elicitation: Application of a new paradigm for job analysis

Sterling Wiggins[1], Jennifer Pagan[2], Michael J. Keeney[1], Amy Bolton[3], Webb Stacy[1], Jeffrey M. Beaubien[1,]

[1]Aptima, Inc.
Woburn, MA 01810
swiggins@aptima.com , mkeeney@aptima.com, wstacy@aptima.com jbeaubien@aptima.com

[2]Naval Air Warfare Center, Training System Division
Orlando, FL 32826
jennifer.pagan@navy.mil

[3]Office of Naval Research
Arlington, VA 22203
amy.bolton@navy.mil

INTRODUCTION

We describe an alternative to traditional self-report job analysis methods that is now available through the increased computing power, networking capabilities, device portability, and display quality that has changed the very nature of work. The methodology we describe integrated self-report data and simulator system-generated data, providing greater insight than either source provides alone and is an enhanced alternative to traditional self-report job analysis methods. This changing nature of work is evidenced in the percentage of jobs that are performed through the use of computers, all of which can potentially be fitted with simulations that exceed what, in the past, had been considered as "high fidelity" (Whetzel, McDaniel, & Pollack, 2012). These advanced simulations can be instrumented and used as interactive tools to design work, to understand how work is performed, and to deliver training to perform work. Our research reports on such an effort. Contextualized knowledge elicitation employs multiple data sources collected simultaneously, in real-time, avoiding issues of method bias (method variance, or common method bias). The results from multiple sources mutually reinforce each other, balancing strengths and weaknesses and form a virtuous circle, a concept that poses a recurring cycle of events, the result of each one being to increase the beneficial effect of the next (Oxford Dictionary, 2010). The combination of system-generated data and self-reports can provide cues to guide investigations, by using self-reports to highlight where to look within voluminous system data, and system data can show where verbal follow-ups are appropriate.

Our method was applied during work to evaluate the potential contributions of MAGIC CARPET, a technology under development to assist Naval Aviators in landing on the deck of an aircraft carrier, one of the most difficult and dangerous tasks that a Naval aviator must perform. The *Maritime Augmented Guidance with Integrated Controls for Carrier Approach and Recovery Precision Enabling Technologies* (MAGIC CARPET) project integrates automated throttle control during final approach with onboard sensors, control capabilities, aircraft displays, and external visual landing aids to increase safety and reliability, and reduce workload. While the current training continuum continues to be highly effective, the total ownership costs associated with live training and the need to maximize aircraft usage beckons for an improvement in aircraft capabilities to reduce initial training and currency requirements.

While touching on the results for MAGIC CARPET, the focus of our report is on an exciting methodology, the contextualized knowledge elicitation workshop, which we used to develop the results we report. The workshop event was conducted in December 2013 at the Manned Flight Simulation operation at Naval Air Station Patuxent River, Maryland. The purposes of the workshop were to gather preliminary data about MAGIC CARPET effectiveness, safety, and training requirements that would provide hypotheses and refine methodology for a subsequent formal experiment. Participants in the workshop used the F/A-18 simulator to land on an aircraft carrier using both conventional technologies and MAGIC CARPET, in varied conditions designed to reduce the mental workload during the landing passes. The simulator recorded the landing event, electrocardiogram (EKG)-based workload measures, eye tracking, and a variety of simulator-based measures. For some participants, an experienced Landing Signal Officer (LSO) was provided to serve in that role in an LSO simulator which was virtually connected to the FA-18 simulator. Contextualized knowledge elicitation occurred after each landing. The simulator replayed the landing that the participant had just completed while the participant remained seated in the cockpit and described his thought process during the landing exercise. The replays could be paused to allow researchers to conduct detailed discussions with the participant.

SHORTCOMINGS OF TRADITIONAL JOB ANALYSIS METHODS

Current approaches to data collection in job analytic methodology (for example, Fine & Cronshaw, 1999; Wilson, Bennett, Gibson, & Alliger, 2012) rely on fallible human memory, understanding of researcher questions, and accurate researcher recording and interpretation of participant responses. This is because traditional job analytic methods all depend ultimately upon the same principal data sources, retrospective self-reports from job incumbents and their managers or supervisors, collected through surveys, interviews, and focus groups, or observations by trained human observers. Researchers therefore depend on the ability of incumbents, often in surroundings very different from the actual work, to understand what information is being sought, perform accurate search and retrieval from memory of the desired information, and then accurately describe their memories. Each step of this process offers opportunity for misunderstanding, memory errors, and mistakes in recording participant responses or psychometric and methodological issues that are documented throughout the research literature (for example, Rogelberg, 2002; Wilson, et al., 2012). Although work observations can occur while actual work is being performed, much job analysis occurs away from the environmental cues present in hazardous or confined locations, during infrequent or unpredictable tasks (such as response to an actual emergency), or interactions with technology. Absent the contextual cues normally present in noisy, stressful, and dangerous environments, participants may not be able to recall and provide insight into all aspects of task performance.

Expert performers are typically desirable as job analysis participants, but expertise inherently involves the development of highly automatized processing that occurs outside conscious awareness. Reports from participants, regardless of data collection method, can also be influenced through participant efforts to manage the impression and representation of themselves they present to the researcher (for example, see Jansen, König, Kleinmann, & Melchers, 2012). When all data are collected from the same source at the same time, method bias, a source of systematic error, can artificially enhance or attenuate statistical relationships and differences (Conway, 2002; Conway & Lance, 2010). Better understanding of work requirements and processes are available through enhancements in simulation technologies and conceptualizations of *simulation fidelity*, meaning an accurate determination of characteristics and presentation accuracy and the degree to which the simulation accurately mimics characteristics, features, movement, and responses relevant to the work and investigatory goals involved (Stacy, Walwanis, Wiggins, & Bolton, 2013). Simulations can consistently replicate representations of work situations, record and assess participant interactions with these environments in real-time, without the distraction of interacting with researchers, and replay these interactions to the participant and researchers for further investigation using cognitive task analysis methods (Crandall, Klein, & Hoffman, 2006; Rosen, Salas, Lazzara, & Lyons, 2012).

After 50 years of usage, we propose conventional job analysis methods have reached their ultimate development, and may not be optimal for modern tasks that depend heavily on cognitive skills, particularly perceptual-cognitive skills, rather than observable physical behaviors. This is because despite their indisputable value, the industry standard approaches to job analysis currently in use suffer from a number of shortfalls. They depend on reports or observations of behaviors of job incumbents. Since the earliest efforts to formally research people at work, researchers have known that the simple act of studying something can change it (Greenwood, 1986; Hassard, 2012; Sonnenfeld, 1985). These methods are also labor intensive, and can be especially burdensome on research participants. Most critically, they are limited in their ability to provide insight into task performance that is often not available to self-report or by observations.

THE VIRTUOUS CIRCLE AND CONTEXTUALIZED KNOWLEDGE ELICITATION

In contrast to conventional job analysis methods, contextualized knowledge elicitation occurs within the context of work performance by using high-fidelity simulations of the work. These simulations enable multiple incumbents to be presented with a controlled consistency in stimuli, and enable researchers to pause and replay stimuli and participant actions while answering questions and providing explanations. This situation helps to ensure that thought processes most relevant to the task are activated and memories are fresh and foremost in the participant's mind. Further, data collection is not limited to accurate self-reports, because the participant's actions in the simulation had previously been recorded by unobtrusive instrumentation. As a result, data from system-based measures, which are much harder for participants to influence than off-line self-reports, can be used to provide an objective corroboration of participants' descriptions, explanations, and comments. This combination of system-based and highly contextualized self-reports minimizes the potential that common method variance may distort research conclusions.

SYSTEM-COLLECTED DATA

The sources for system-collected data are sensors and measurement devices that are built into the simulator or attached to the participant. The key difference between system-collected and self-report data is that system-collected data (a) can document participant behaviors and reactions at intervals that are as short as the researchers wish to deal with (for example, periods of milliseconds to seconds or minutes), (b) depend upon behaviors and responses, not fallible human memory, (c) are largely outside the ability of participants to manipulate, other than by altering their own behaviors or reactions. System-collected data can measure a variety of variables. The first category of measures document participant actions and behaviors through positions of controls (documenting settings, adjustments, and inputs across time) and voice and text communications. A second category of system-generated measures assess participant behaviors that cannot be recorded through the simulator, such as the tracking of eye movements. A third category of system-generated measures measure participant physiological conditions, providing direct or indirect insight into variables such as workload and stress. In our research, we collected a variety of data through systems embedded in the simulation technology. The first set of data described pilot performance as reflected, in this application, by control inputs and their effect on the position and flight path of a simulated aircraft as it approaches a carrier to land. We also used eye-tracking instruments to assess where the pilot was looking and EKG data to compute estimates of workload. During this research, we identified an unforeseen additional capability of system-based measures. This was their capability to assess effects of pilot behaviors that cause accelerated deterioration (wear-and-tear) of equipment through rough usage, abuse and other behaviors such as hard landings and high braking levels. These behaviors can be difficult for observers to spot or be willing to report and are unlikely to appear in self-reports.

MANUALLY COLLECTED AND SELF-REPORT DATA

We also collected a set of manual and self-report measures. For passes that had LSO support, LSO scores were recorded manually, using their standard criteria and measures (such as those in NAVAIR 00-80T-104). After each landing, a researcher collected subjective workload ratings from the participant, using the Bedford scale (Corwin et al., 1989; Gawron, 2000), a one-dimensional self-report rating by the participant of his or her spare mental capacity while completing a task. This measure is widely used in aviation because it is simple, quick, and easy-to-use in high-workload environments.

PARTICIPANTS

The eight participants represented a variety of levels of training experience, from highly experienced F/A-18 pilots to having never before flown an aircraft. Our primary participants were six qualified Naval Aviators. In addition, we had four LSOs (who were qualified aviators), two of whom also participated as pilots. Two non-pilot engineers, with no previous flying experience, provided the ability to estimate the effects of MAGIC CARPET independent of flight qualifications. A total of eight participants flew the simulator. In the remainder of this report, we use *participant* to refer only to the eight individuals who flew the simulator, and do not include the two individuals who served only as LSOs.

LSOs are Naval aviators with additional training to enable them to facilitate safe and expeditious recovery of aircraft aboard a carrier. LSOs inform landing pilots about known environmental conditions, acknowledge pilots as they begin their final descent, talk the pilots in when the pilots have a problem perceiving their normal landing cues, and tell pilots not to land when conditions exceed safety limitations. They also annotate and score every pass made by a pilot during carrier qualification (CQ) landings. The usual term used for their activities is *waving* pilots in (although they no longer do so, in the past LSOs used signal paddles and body position to provide information to landing pilots, hence they are often referred to as *Paddles*).

Participants represented various degrees of F/A-18 carrier landing experience. The two non-pilot engineers were generally familiar with aviation, but had no flying experience and therefore no landing experience of any kind. Three Naval aviators were Fleet Replacement Squadron (FRS) students, having no experience in landing an F/A-18. One pilot had been on a single tour, but had fewer than 50 *traps* (a trap is a single arrested carrier landing), and the remaining two Naval aviators were very experienced, each having over 300 traps. We grouped the FRS students and the pilot with a single tour into a category labeled *novice*. The two qualified pilot-LSO aviators were grouped as experts, and the two non-pilot participants were labelled *Ab Initio*.[1]

CONDITIONS

We used three sets of landing conditions. The first reflected normal landings, a second set of conditions used a dual tasking to assess the degree to which MAGIC CARPET ran the risk of automation-induced complacency, and a third turned off the simulator's heads-up display (HUD), which Naval aviators normally depend upon during carrier landings. This final condition yielded insights into how dependent pilots would be on the new HUD configuration and into the effect that MAGIC CARPET would have on their landings in off-nominal conditions. Two expert pilots, who at first expressed skepticism of MAGIC CARPET, attempted a landing with the aviator blindfolded, while the other served as LSO. This landing, which would have been extremely difficult using conventional technology, was successful.

METHOD

Participants were briefed on the purpose of the study and then participated in two calibration tasks prior to flying in the simulator. The first calibration task performed a Multi-Attribute Task Battery (MATB) test to establish three levels of baseline workload (low, medium, high). EKG data were also collected. MATB data were used to calibrate the real-time workload assessment measurements. The second calibration task seated participants in the flight simulator while an experimenter assisted them in capturing eye gaze poses to calibrate the eye tracker.

Participants were next given an opportunity for free-flight in the simulator to familiarize themselves with its characteristics. To reduce variability in behavior of non-pilot participants, the *autothrottles* (ATC) system, which automatically maintains the aircraft in the correct airspeed and angle of attack (AoA) during landings, was engaged during their passes. However, to elicit the most natural behavior from pilots, they were asked to use the ATC setting they normally use. All pilots preferred ATC to be disengaged for passes that used conventional technology. For MAGIC CARPET passes, ATC was always engaged because the ATC is integrated into the MAGIC CARPET technology.

After the practice session, which generally lasted 10 minutes, participants landed using conventional landing technology. Non-pilot engineers were given a single opportunity to land under nominal conditions. Other participants were given two or three opportunities to land, one with ideal winds (25 knots straight down the landing area, called "down the angle[2]") and another with somewhat more difficult but still typical wind conditions (30 knots down the axis of the carrier, which is 9 degrees off from "down the angle.")

Participants were asked to land in two dual-task conditions. The first dual task asked them to monitor a colored light located outside of their foveal vision. The light could have one of four colors: white, amber, red, or green, and participants were asked to respond only when the light turned amber. This task required participants to look to the left about 30 degrees to see the light, and press the trigger when the light turned amber. Participants were instructed to pay attention to landing as their primary task and to respond to the secondary task only as they could. This condition was expected to add to the landing workload.

Participants then completed a brief on MAGIC CARPET and were given an in-simulator demonstration and an opportunity for free-flight using MAGIC CARPET. ATC was engaged for all participants for all MAGIC CARPET passes. In these passes, the participants repeated the landing and light monitoring tasks. In addition, for the MAGIC CARPET portion of the session, a second dual-task introduced an inattentional blindness condition, in which one or more unexpected objects were introduced into the scene. Finally, pilot participants were asked to land without a HUD, first using MAGIC CARPET technology, then using conventional technology.

After each pass the researchers displayed a playback of the just-completed landing performance to the participant. Knowledge elicitation occurred during the replay, when researchers asked the participants to pause and describe important shifts in their attention, perceptions of fluctuations in the workload they experienced, and key impressions

---

[1] A Latin term meaning "from the beginning."

[2] The landing area of US carriers is offset 9 degrees from the centerline of the ship.

about the pass. For job analysis, the critical issues were that (a) knowledge elicitation occurred in the full context of the environment in which the work is performed in this application, an aircraft cockpit; (b) immediately after the job-relevant behaviors were performed and recorded; and (c) about the recordings, using the specific features displayed in the recordings as stimuli for further questions.

Analyses were designed to investigate three aspects of the MAGIC CARPET technology, which are displayed in Table 1.

Table 1. Research Questions

Effect	*Research question*
Main effects of technology on landing quality.	How did landings using MAGIC CARPET compare to landings using conventional technologies overall, averaged across pass types and experience levels?
Effects of technology by experience.	Did MAGIC CARPET affect novices and experts differentially, averaged across pass types?
Effects of technology by experience and pass types.	Did MAGIC CARPET affect novices and experts differentially by pass type?

Three important measures of landing quality are pilot errors in (a) glideslope, (b) lineup, and (c) AoA. The variables were measured by and provided from the Controls Analysis and Simulation Test Loop Environment (CASTLE) aircraft simulation. We present a summary of overall results using bar charts with error bars. The variables are reported as the mean of absolute errors of the three types. The reason for using absolute error is that it is the size of these errors, and not their direction, that is of most interest when assessing overall differences. In fact, averaging over raw errors that could be either negative or positive runs the risk of cancelling out meaningful data. Absolute errors do not have this liability. That said, the workshop was not an experiment, and the results cannot be interpreted in the way that it would be in an experiment. The workshop was primarily a knowledge elicitation exercise intended to help formulate dependent and independent variables for a later formal experiment; it lacked the experimental controls that would ordinarily be applied and was not designed with experimental power in mind. Thus, a "significant" result must be interpreted to be a very promising—but not generalizable—result. For this reason, no other statistical inference was performed on these data.

RESULTS

When using conventional technology to fly the first of the two nominal passes, all of the pilots reported mainly looking at the HUD, the vertical rate of descent indicator (VSI), and the left digital display indicator (DDI) early in the pass to scan for the glideslope, and then shifting their attention to the ball around 1.5 nautical miles (NM) behind the ship. All of the pilots reported how easy it was to fly the pass using MAGIC CARPET. One of the expert pilots said, "*I thought that was pretty easy. I thought that [MAGIC CARPET] was awesome. It's almost so easy it puts you to sleep. The ball is staying right where I left it at and that's awesome to see.*" In fact, the expert pilots later cited this lack of workload as a concern, stating that it might induce automation complacency.

Several of the novice pilots reported experiencing scan breakdown when flying the second, more challenging nominal pass that contained axial winds. One novice pilot described his experience flying this pass as, "*I was fighting lineup the whole time. I was not watching VSI. I was spotting the deck.*" Even one of the most expert pilots said, "*I made a noticeable number of corrections because of the axial winds.*" Pilots also conveyed that the level of workload was consistently higher throughout the pass when compared to the first pass with winds coming down the angle. Even though this was a fairly challenging pass for some pilots, they all reported that it was much easier to fly when using MAGIC CARPET. In fact, the expert pilot who mentioned making a number of corrections when using the conventional technology, said after flying the same pass using MAGIC CARPET, "*It was easy to handle the axial winds. The winds appeared to be almost negligible because it's so easy to keep it [the jet] on centerline. It didn't seem like axial winds because it was so damn easy to keep it on centerline. Deviations were smaller and the corrections were easier. I made a couple of small deviations to make sure of where I want to be.*" Using MAGIC CARPET, all of the pilots reported a reduction in workload as they did not experience the shifts in attention on this pass they experienced using conventional technology. The self-report about the nominal passes are substantiated by the system-collected data shown in *Figure 1* and *Figure 2*.

It can be seen in *Figure 1*, that MAGIC CARPET showed a significant advantage for all participants in lineup and glideslope error, and that there was little if any difference for AoA error. The most likely explanation for this is that AoA error was already small with conventional technology, so there was not much room to improve. Figure 2 shows the dramatic advantages in lineup and glideslope error. Note the 0.4 NM burble "bump" in all charts that reflects the burble (the turbulent air behind the carrier due to its motion through the air). These included a pass with winds at 25 knots down the angle and another at 30 knots axial. As in other charts, difference in AoA error was effectively zero.

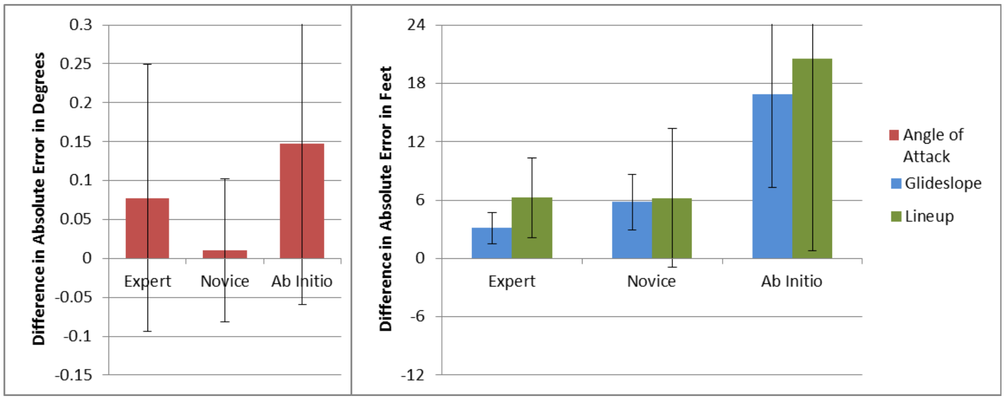

Figure 1. Difference of absolute error between conventional and MAGIC CARPET technologies for glideslope, lineup, and AoA for nominal passes showed a significant or almost-significant advantage for the MAGIC CARPET landings in lineup and glideslope error.

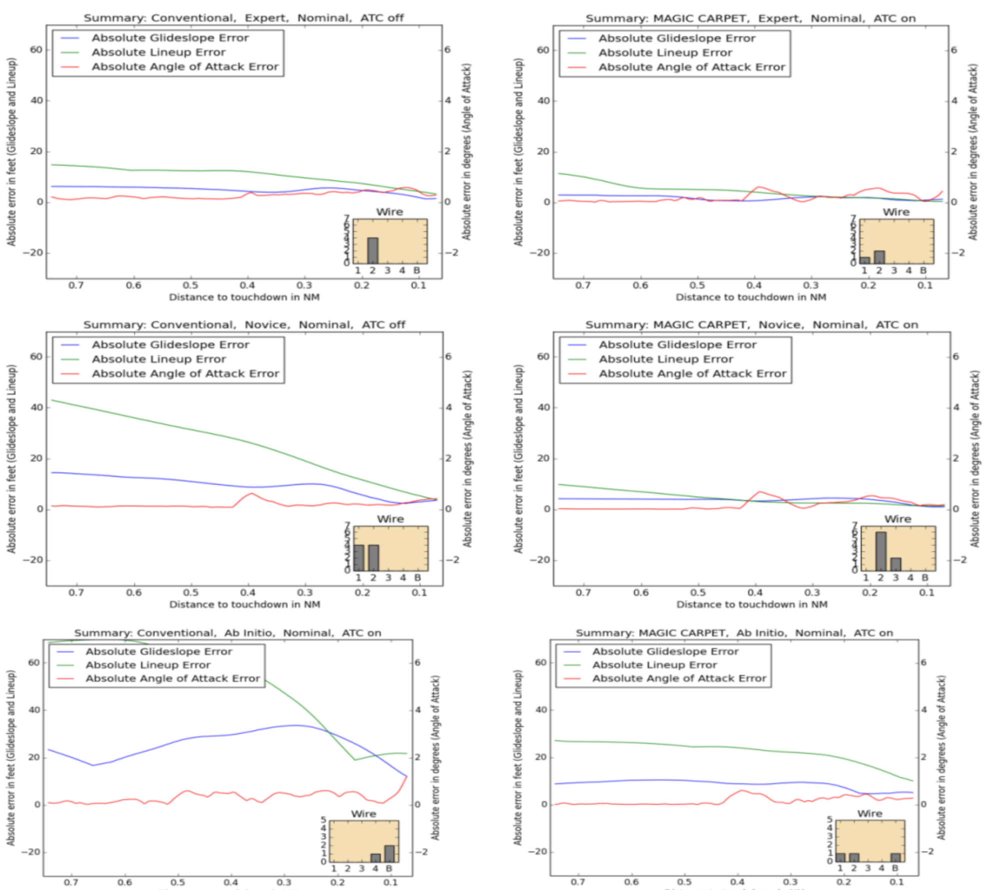

Figure 2. Absolute glideslope, lineup, and AoA error for conventional and MAGIC CARPET technology.

Figure 2 also shows that experts were already doing very well on nominal passes, so there was less room to show the advantages of MAGIC CARPET. For Novice and Ab Initio participants, however, there were large advantages in the MAGIC CARPET passes, which are visually very evident in the charts.

Note that MAGIC CARPET appeared to reduce novice errors nearly to the level of expert errors using MAGIC CARPET (and below the level of expert errors using conventional technology.) This supports the assertion we heard from LSO during the workshop that *"MAGIC CARPET makes novices look like Top-10 ball flyers*." These charts support the subjective reports from the pilots, that MAGIC CARPET enabled smoother and better landings, though this was somewhat smaller for the experts than for the others.

Figure 3 shows that participants, in general, experienced less workload during the nominal passes when using MAGIC CARPET, as compared to conventional technology.

Two dual-task conditions were used during the workshop, spare capacity and inattentional blindness. The first of the dual- task conditions required participants to look to the left about 30 degrees to see a light that illuminated periodically, whose color changed over time, and press the trigger when the light was amber. Participants were instructed to pay attention to the primary task (landing) and to respond to the secondary task when they could (looking for the amber light). The purpose of this condition and the second dual-task was to provide insight about the degree to which MAGIC CARPET enables pilots to pay attention to stimuli that would otherwise go unnoticed, and to gain some preliminary insight into whether MAGIC CARPET might induce automation complacency. When flying this condition using conventional technology, pilots had intermittent success identifying the amber light before calling the ball. Soon after calling the ball, they shifted their attention to the landing task and noticed fewer of the amber lights. Pilots reported that when they tried to pay attention to looking for the amber light after calling the ball, they often lost line-up or deviated from glideslope, often making large corrections to recover. As one pilot noted, "*I ended up pushing it pretty high when looking for the lights and then had to make an aggressive correction.*" Most pilots said they gave up trying to identify the amber light around 1 mile from the ship, with the most experienced pilots giving up at 0.7 miles.

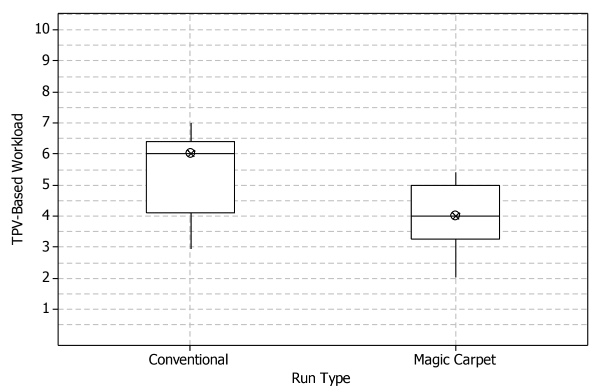

Figure 3. TPV-Based Workload for nominal passes for all participants as a function of pass type.

When pilots used MAGIC CARPET to fly this condition, they reported having more time to look for the amber light and could still fly a good pass. One pilot used a strategy to look for the lights every 1 to 2 seconds, and said he could do this because "*I could look away and the flight path would be where I left it. It would be a lot more predictable where the ball would be using MAGIC CARPET.*" Because the pilot did not have to shift his attention between the HUD, making corrections, assessing the corrections, and then taking a quick look for the amber light, it allowed him to spend more time looking for the light. *Figure 4* shows the familiar MAGIC CARPET advantage for glideslope and lineup error, which were significant or almost-significant. Both experts and novices showed a glideslope advantage comparable to the nominal passes, and the novice lineup advantage was comparable to the nominal passes. However, unlike the nominal passes, experts showed a large and significant performance improvement under MAGIC CARPET for AoA error. AoA error was significantly reduced for experts and glideslope and lineup error was significantly or almost-significantly reduced for both experts and novices. This is another example of how the contextualized knowledge engineering approach produces a virtuous circle: one set of results (the system-collected data) increases the beneficial effect of another set of results (the self-report data) by providing objective corroboration of participants' descriptions, explanations, and comments.

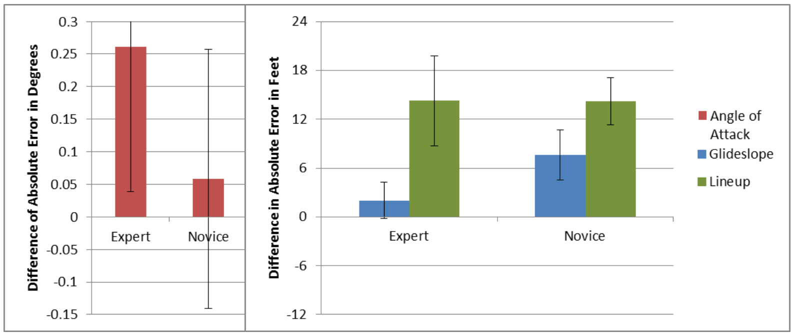

Figure 4. Difference in absolute error between conventional and MAGIC CARPET landings for dual task passes.

DISCUSSION

As with the other conditions, the evidence suggested a clear advantage for MAGIC CARPET over conventional technology for the landing task. This use of playback of real-time data, reflecting natural work performance, enables researchers to construct probes for additional information about focus of attention, change of attention, workload changes, and decision-making. It enables clarification and detailed probes of key points that could not be observed or recorded through any of our other measures. The combination of system-collected and self-report data allows mutual reinforcement and cross-cuing, in which system-collected data suggest topics for interviewing and cognitive task analysis (CTA) follow-up, while the self-report results can indicate how to refine and where to focus analysis within the system-collected data. Finally, although data from a single participant were collected during a single session, we collected at least two independent data sources, self-reports and system-based data, and when possible a third from the LSO reports, reducing the potential that our results were altered by method bias.

The virtuous circle and contextualized knowledge elicitation approach provides insight into task performance that cannot be obtained through observation or self-reports or system-data alone, such as strategies to adjust work to changing situations and conditions. This approach overcomes many challenges participants encounter in recognizing, recalling, describing, and reporting highly-automatized behaviors, particularly when their reports can be interactive combined effects of high workload, high stress, and highly automatized behaviors. It provides the ability to present known stimuli to participants rather than asking them to imagine, reducing the potential for results to be altered because participants understood or imagined different situations. Traditional approaches using incident-based interviewing techniques and survey items ask participants to recall incidents and hypothesize about situations, with no assurance that participants are recalling incidents that reflect the task demands and environmental conditions of greatest interest to the researcher.

This approach also enables researchers to alter conditions to assess behavior in real-time by manipulating situations and measuring resulting changes in participant behaviors and responses. In contrast, our approach places participants in the full context in which the work is performed, allows participants to work without interruption, and then inquires about decisions and perceptions, using unambiguous recordings of the participant's task performance, immediately after performance while the memories are still fresh to the participant. Further, researchers can use system-recorded data to craft verbal follow-ups, and then use the results of verbal inquiries to further refine system-based data collection. The virtuous circle and contextualized knowledge elicitation approach provides greater ability to objectively and accurately measure differences, such as we did by replicating performance effects due to equipment failures (HUD outage). Researchers can manipulate attentional workload by adding secondary tasks and distractions, and then observe effects on performance of primary task. Researchers can collect accurate behavioral data, present consistent stimuli across participants, without regard to expertise, verbal ability, and past real-world experiences. The advantages of this approach make it a good candidate for use in designing training scenarios, developing measures for training effectiveness, and conducting formative evaluations of system designs.

ACKNOWLEDGEMENTS

This work was performed under contracts N0001413P1217 and N0001412C0058. We would like to thank the Office of Naval Research (Code 35 and Code 34) for their support of this activity, and others on the research team, including those at the Navy's Manned Flight Simulator, at the Naval Air Warfare Center Aeronautics Division, at the Naval Air Warfare Center Training Systems Division, and at the University of Iowa's Operator Performance Laboratory for their valuable discussions and comments. The views expressed herein are those of the authors and do not necessarily reflect the official position of the organizations with which they are affiliated.

REFERENCES

Conway, J. M., & Lance, C. E. (2010). What reviewers should expect from authors regarding common method bias in organizational research. *Journal of Business and Psychology, 25*, 325-334.

Conway, J. M. (2002). Method variance and method bias in Industrial and Organizational psychology. In S. G. Rogelberg (Ed.) *Handbook of research methods in Industrial and Organizational psychology* pp. 344-365. Malden MA: Blackwell.

Corwin, W. H., Sandry-Garza, D. L., Biferno, M. H., Boucek, G. P., Logan, A. L., Jonsson, J. E., & Metalis, S. A. (1989). *Assessment of crew workload measurement methods, techniques, and procedures. Volume I - Process, methods, and results* (WRDC-TR-89-7006). Wright Patterson Air Force Base, OH: Air Force Systems Command.

Crandall, B., Klein, G., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to Cognitive Task Analysis*. Cambridge, MA: The MIT Press.

Fine, S. A., & Cronshaw, S. F. (1999). *Functional job analysis: A foundation for human resources management*. Mahwah, NJ: Erlbaum.

Gawron, V. (2000) *Human Performance Measures Handbook*. Mahwah, NJ: Erlbaum.

Greenwood, R. G. (1986, August). The Hawthorne studies. *Academy of Management Best Papers Proceedings, Supplement*, 24-35. http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=c2f23252-3589-4f56-b694-d053a636f083%40sessionmgr4001&vid=1&hid=4112

Hassard, J. S. (2012). Rethinking the Hawthorne Studies: The Western Electric research in its social, political and historical context. *Human Relations, 65*, 1431-1461.

Jansen, A., König, C. J., Kleinmann, M., & Melchers, K. G. (2012). The interactive effect of impression motivation and cognitive schema on self-presentation in a personality inventory. *Journal of Applied Social Psychology, 42*, 1932-1957.

NAVAIR 00-80T-104 (2001, December). *NATOPS Landing Signal Officer Manual*. Washington, DC: Naval Air Systems Command. Available at http://www.navyair.com/LSO_NATOPS_Manual.pdf .

Oxford Dictionary of English (3rd Edition) (2010). Oxford University Press.

Rogelberg, S. G. (2002). *Handbook of research methods in Industrial and Organizational psychology*. Malden MA: Blackwell.

Rosen, M. A., Salas, E., Lazzara, E. H., & Lyons, R. (2012). Cognitive task analysis: Methods for capturing and leveraging expertise in the workplace. In M. A. Wilson, W. Bennett, S. G. Gibson, & G. M. Alliger (Eds.) *The handbook of work analysis: Methods, systems, applications, and science of work measurement in organizations* (pp. 185-200). New York: Routledge.

Sonnenfeld, J. A. (1985). Shedding light on the Hawthorne studies. *Journal of Occupational Behavior, 6*, 111-130.

Stacy, W., Walwanis, M., Wiggins, S., & Bolton, A. (2013). Layered fidelity: An approach to characterizing training environments. In *The Interservice/Industry Training, Simulation, & Education Conference (I/ITSEC).* National Training Systems Association, 2013.

Whetzel, D. L., McDaniel, M. A., & Pollack, J. M. (2012). Work simulations. In M. A. Wilson, W. Bennett, Jr., S. G. Gibson, & G. M. Alliger (Eds.), *Series in applied psychology. The handbook of work analysis: Methods, systems, applications and science of work measurement in organizations* (pp. 401-418). New York: Routledge/Taylor & Francis Group.

Wilson, M. A., Bennett, W., Gibson, S. G., & Alliger, G. M. (Eds, 2012). *The handbook of work analysis: Methods, systems, applications and science of work measurement in organizations*. New York: Routledge.