Air Superiority Knowledge Assessment System

Sara Elizabeth Gehr
Boeing
Mesa, AZ
Liz.Gehr@mesa.afmc.af.mil

Christopher C. Metz
Rickard Consulting Group, Inc.
Mesa, AZ
Christopher.Metz@mesa.afmc.af.mil

Brian T. Schreiber
Lumir Research Institute
Mesa, AZ
Brian.Schreiber@mesa.afmc.af.mil

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

ABSTRACT

The US Air Force is developing a multifaceted approach to assessing knowledge and measuring mission effectiveness and performance in military training and rehearsal environments. This paper will present recent results from a field study using new knowledge assessment technology - the Air Superiority Knowledge Assessment System (ASKAS). ASKAS is a system that has been developed to enable measurement of a pilot's mastery of Mission Essential Competencies (MECs^{SM}). The pilots watch a series of mission oriented combat engagements which are actual air combat scenarios recorded from simulators, including audio communication between the pilots and the controller. After approximately two minutes of playback, the engagement stops, and the pilot is asked a series of questions. The questions are presented first, followed by the responses, allowing response times to be measured. The questions assess the pilot's knowledge about the events of the preceding vignette, specifically addressing the MECs^{SM} that are necessary for a successful mission. This process is repeated until the end of the engagement. The paper will present recent results, which include both percent correct and reaction time as a measure of the pilot's situational knowledge and mastery of the MECs^{SM}. Individual scores are compared to other pilot scores with different amounts of experience; more experienced pilots should have higher scores. We will discuss operational implementation challenges with ASKAS, the assessment protocol developed for field use, and possible future uses for this tool. For example, if a pilot takes the test before and after training events such as Distributed Mission Operations (DMO), training effectiveness can be measured, and will be indicated by higher scores after training. This knowledge assessment system could also be used to pinpoint a pilot's knowledge gaps and outline areas that may need more attention in training.

ABOUT THE AUTHORS

Sara Elizabeth Gehr is a Human Factors Design Specialist with Boeing at the Air Force Research Laboratory, Warfighter Readiness Research Division, in Mesa, AZ. She received her Ph.D. in Experimental Psychology from Washington University in St. Louis in 2001.

Brian T. Schreiber is CEO Senior Scientist with Lumir Research Institute at the Air Force Research Laboratory, Warfighter Readiness Research Division, in Mesa, AZ. He completed his M.S. in Human Factors Engineering at the University of Illinois at Champaign-Urbana in 1995.

Christopher C. Metz is an F-16 Subject Matter Expert at the Air Force Research Laboratory, Warfighter Readiness Research Division, Mesa, AZ. He has over 3000 hours in fighter aircraft to include 1500 hours in the F-16. He received a B.S. in Engineering Mechanics from the United States Air Force Academy in 1982 and his M.S. in Business Organizational Management from the University of La Verne in 1995.

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 from Texas A&M University in 1995.

Air Superiority Knowledge Assessment System

Sara Elizabeth Gehr
Boeing
Mesa, AZ
Liz.Gehr@mesa.afmc.af.mil

Christopher C. Metz
Rickard Consulting Group, Inc.
Mesa, AZ
Christopher.Metz@mesa.afmc.af.mil

Brian T. Schreiber
Lumir Research Institute
Mesa, AZ
Brian.Schreiber@mesa.afmc.af.mil

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

INTRODUCTION

Evaluating the performance specific combat skill proficiency and current knowledge of a combat pilot during a training exercise is a difficult process. The situations are dynamic, involve interactions among multiple players, and aside from simple outcome measures, established measures of knowledge and SA do not exist. Many researchers have attempted to measure pilot performance in training environments, both in Distributed Mission Operations (DMO) and live-fly events. These methods often focus on subjective ratings of performance by Subject Matter Experts (SMEs) (Krusmark & Schreiber, 2004, Crane, Robbins & Bennett 2000) or, more recently, objective measures of performance (Schreiber, Watz and Bennett, 2003; Schreiber, Watz, Bennett & Portrey, 2003). Even when multiple measures of pilot performance are used to evaluate training (Gehr, Schreiber & Bennett 2004), these methods usually evaluate outcome and process measures, but are unable to directly measure the amount of knowledge a pilot has acquired, or how quickly and effectively a pilot can acquire that knowledge.

The introduction of a knowledge assessment test would enable further validation of DMO or live-fly training. If administered before and after training, in conjunction with current outcome and process measures, both subjective and objective, it would provide further evidence for the value of that training. It will provide information about how well that training has helped the pilot learn how to more effectively acquire knowledge about a current situation. This is not currently being measured in most training scenarios. In addition, different methods of training, or different syllabi, could be compared in their ability to help a pilot learn to acquire knowledge.

Further motivation for developing an objective measure of acquired knowledge is anecdotal evidence that pilot experience, often measured by flight hours, years of experience, or rating, may not be the best measure of a pilot's ability to acquire knowledge about a complex combat situation. Because of quickly changing technology, more recently trained pilots may have a better understanding of current tactics and other current skills that give them a better ability to acquire knowledge than more experienced pilots. The more experienced pilots may be not be used to the new technologies, and may take longer to learn them, due to interference with previously learned methods, than a new pilot who has no experience with other methods. In addition, the less experienced pilot may have had more recent training than the more experienced pilot which may lead to better performance. A knowledge acquisition test would allow evaluation of a pilot's ability to gather and integrate knowledge, unbiased by his years of experience.

The Air Superiority Knowledge Assessment System (ASKAS) was developed to measure a pilot's ability to acquire and integrate knowledge from an air-to-air combat scenario. The system is based on Mission Essential Competencies (MECs). MECs are defined as "Higher-order individual, team, and inter-team competencies that a fully prepared pilot, crew or flight requires for successful mission completion under adverse conditions and in a non-permissive environment" (p. 12-2, Colegrove & Alliger, 2002). The ASKAS methodology evaluates a pilot's mastery of the knowledge and skills identified in the MEC process. These are assessed within a combat relevant context, a key component of the MEC process, as an engagement unfolds. ASKAS can evaluate how well the pilot has acquired knowledge about each MEC from the engagement, and, eventually, which MECs he has mastered or not.

The ASKAS development approach is driven by research on the development and validation of Situational Judgment Inventories (SJIs) and Job Knowledge Inventories (JKIs) (see Hanson & Borman, 1993; Hedge, Hanson, Borman, Bruskiewicz & Logan, 1996). These inventories have been developed and validated in a variety of complex domains where more traditional knowledge assessment tools have not proved to be adequate for the task. In addition, SJIs and JKIs were recently shown to have substantial incremental validity as predictors of job performance (Clevenger, Pereira, Wiechmann, Schmitt, & Schmidt Harvey, 2001).

SJIs are more context- or situational-based assessments of performance. The traditional way SJIs work is that a respondent is presented with a written description of a job-relevant situation. Once they have read the situation they are asked to respond to a set of possible responses which are also presented in written format (Paullin, McKee, Hanson, & Hedge, 1994). More recently there have been successful applications of SJIs using videotaped presentations of the situation followed by a set of questions.

JKIs are tests that require individuals to answer multiple choice questions related to critical aspects of their on-the-job knowledge, skills, and abilities. They have been shown to be particularly useful for assessing proficiency related to job technical information and as criterion measures. When properly developed, these inventories representatively sample the domain of interest and the level of knowledge a given individual has relative the various aspects of the work domain (Paullin, McKee, Houston, Hanson, & Hedge, 1997). ASKAS represents a first attempt to assess the feasibility of using this approach for assessment in a complex air combat domain.

The ASKAS methodology is also somewhat similar to methods of measuring SA (e.g., Endsley, Sollenberger, & Stein, 2000). For example, SAGAT (Endsley, Selcon, Hardimon, & Croft, 1998) freezes a simulation scenario at random times while an operator (pilot, air traffic controller, etc.) is engaged. It then asks the operator questions about what was happening at the time of the freeze. While similar in many regards, ASKAS, through the incorporation of MECs, has the capacity to tap specific skill proficiency and underlying knowledge across the range of performance expected in combat scenarios. This permits ASKAS to provide information identifying specific areas of weakness.

This paper discusses developmental aspects of the ASKAS methodology and the results of a preliminary study of its application as an assessment tool for knowledge and skill in a small group of current F-16 pilots.

METHODS

The Air Superiority Knowledge Assessment System (ASKAS) was designed to assess an F-16 pilot's knowledge about an air-to-air combat scenario by having a pilot watch a pre-recorded engagement of a 4 v 4 or 4 v 6 scenario. After completing a demographic questionnaire, the pilot is presented with some background and instructions about ASKAS on the computer. ASKAS is self-paced and the pilot passes through several screens that introduce the participant to the methodology and the computer-based assessment process. The participant is given information about the engagement he is about to watch. When the pilot has finished reading the background and brief information, the pilot clicks to start the ASKAS program.

When the engagement starts, the pilot can see the four information displays (Radar Warning Receiver (RWR), Heads up Display (HUD), Radar display, and Tactical Awareness Display (TAD) without datalink) of an F16 simulation from the perspective of of Viper One, the four ship flight lead. The participant also hears all the communications between all the pilots and the AWACS controller. See Figure 1 for a screen shot from ASKAS. The pilot watches and listens for approximately two minutes, when the vignette stops. The participant reads the first of approximately 15 questions, uses the mouse to click on a button to bring up four possible responses, and then selects the most appropriate answer as quickly as possible. After the response, the answers go away, and the next question in the set appears. This continues until the participant has answered all the questions in the set about the first vignette. At that point, the engagement resumes where it left off, and the participant again watches and listens for approximately two minutes, until another set of questions appears. This procedure – watching a vignette from an engagement and answering questions – is repeated two or three times until the end of the engagement. Three engagements are presented.

Figure 1. ASKAS Screen shot

The ASKAS questions were designed to assess MEC knowledge. As such, each question was linked to a particular MEC. Responses to the questions varied in degree of "correctness," with one response being correct, and the others were either completely incorrect, or partially correct based on SME judgments. ASKAS records several measures as the participant proceeds through the program. Chronologically, the program records the amount of time the participant takes to read each of the introductory screens. Then, as the participant reads a question, the question reading time is captured (ended by the click to view possible answers). Next, as the participant views the responses and selects an answer, the program captures response time and records whether or not the answer that the participant selected was correct. If the participant selects the wrong answer, it records which one was selected, and how correct that answer was.

Participants

38 participants from two squadrons at Luke AFB in Glendale, AZ completed ASKAS. The participants were all Instructor Pilot (IP) qualified, with an average of over 1300 F-16 hours.

RESULTS

Each participant answered a total of 120 questions across three engagements. The percent correct was used as the dependent variable for the analyses. The data were separated into two groups, based on whether the participant had been through DMO training at the Mesa, AZ Air Force Research Laboratory. 15 of the 38 participants had been through the training, the other 23 had not. As shown in Figure 2, participants who had been through DMO training had a higher percent correct (67%) than those who had not been through training (63%). This difference approaches significance (t=-1.991, p=0.054).

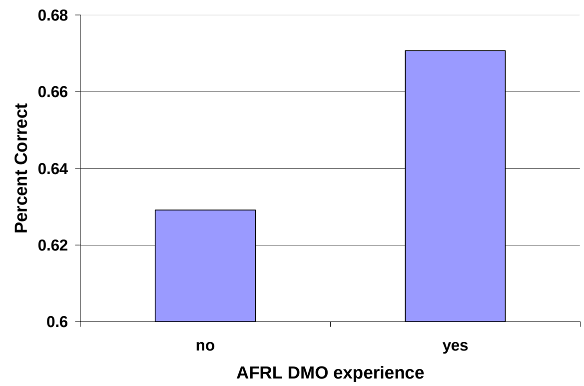

Figure 2. Percent correct as a function of previous AFRL DMO experience

A correlation was calculated between percent correct and experience, as defined by years of service. This correlation was significantly negative (r=-0.518, p=.001). Thus, more experienced participants had lower percent correct than less experienced participants. The scatter plot is shown in Figure 3. Possible reasons for this will be outlined in the discussion.

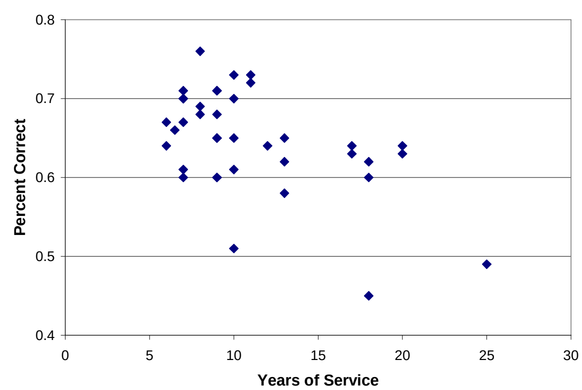

Figure 3. Scatterplot of years of experience versus percent correct

In general, the data for reaction time show non-significant effects. For example, there is no significant correlation between percent correct and reaction time to select an answer (r=-.283, p>0.05). Also, as seen in Figure 4, the correlation between years of service and reaction time at first glance is positive (r=0.456, p = 0.004). However, upon looking at the graph and removing one obvious outlier and rerunning the correlation, this becomes nonsignificant ($r=0.185$, $p>0.05$).

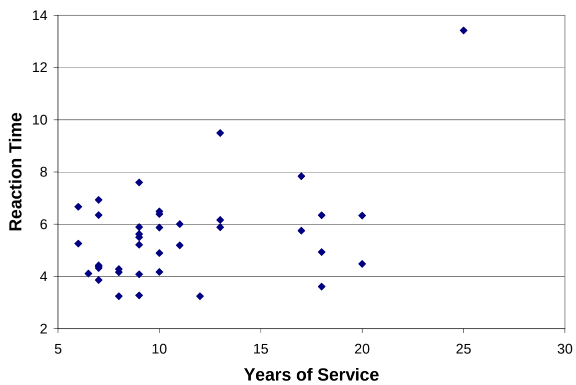

Figure 4. Scatterplot of years of experience versus reaction time

DISCUSSION

This preliminary test of ASKAS shows that it can identify participants with better abilities to acquire knowledge about an air-to-air engagement to which they are watching and listening. Participants who have previously been through DMO training perform at a higher level on this test, indicating that through DMO training they learned techniques to better acquire knowledge about situations. This validates previous research (Gehr et al. 2004) that used different evaluation methods, and also showed DMO training to be effective. Thus, in the future, ASKAS could be incorporated before and after a week of DMO training as another method of assessing the amount of training that occurred during the week. Also, different training syllabi could be compared in their ability to train pilots to more effectively acquire knowledge from an engagement.

The correlation between years of service and percent correct was negative. Although this may seem counterintuitive, after discussing the results with SMEs, we speculate that this actually shows that the traditional measure of a pilot's ability, experience, may not be the most valid measure for a pilot's ability to acquire knowledge from an air-to-air engagement. Another possible explanation is that we have only experienced pilot participant in our sample. This could lead to the observed results in a couple of ways. First, we may have restriction of range, and lower experienced participants may score more poorly than those in our sample, reversing the negative trend. This restriction of range would also explain the nonsignificant correlation of years of service and reaction time. Alternatively, the more experienced pilots may have more duties that do not involve flying for as much of their time. They may have other, administrative, duties that would not give them a chance to fly as much and stay current with the most up to date strategies, doctrine and gameplans. Since these results show that experience may not be a valid measure of ability, in the future thresholds could be set for tests such as ASKAS as a diagnostic tool. Pilots could be required to score above a preset level before they are allowed to progress to the next step in training, or acquire the next higher qualification level.

Another future application for this tool would be in an adaptive training scenario. As mentioned earlier, ASKAS records which skills and knowledge are associated with the questions that the participants answer. In the future, ASKAS could record in which areas the pilot is scoring the lowest. It could then present engagements designed to train those specific skills and knowledge. After the training, ASKAS could retest the pilot with different engagements of similar difficulty. This method would allow pilots to be trained only in areas where they are deficient, and not waste valuable training time on other areas where they are proficient and do not need additional training.

The Air Superiority Knowledge Assessment System is a useful tool for measuring a pilot's current knowledge. Future applications of this tool will enable more precise measurement of increases in knowledge acquisition ability during training, and a more precise measure of a pilot's understanding of a complex combat scenario. It represents a significant advance and extension of both the SJI and JKI methodologies and uses automation for situation or scenario item and knowledge item presentation and for response elicitation and tracking. With our approach to ASKAS, we will link an air combat SJI to an air combat competency-based JKI. The results reported herein underscore the sensitivity of ASKAS as an assessment instrument and a first evaluation of a vignette-based approach for combat-relevant knowledge, skills, and competencies and demonstrate an extremely high fidelity assessment capability that does not exist today.

ACKNOWLEDGEMENTS

The views expressed in this paper are those of the authors and do not necessarily reflect the views or the U.S. Air Force, Lockheed Martin, or Boeing. The authors would also like to thank, in alphabetical order, Deke Joralmon, Peter Neubauer, Karen O'Brien, Toni Portrey, Traci Smith, Eric Watz, and Jack Wooster.

REFERENCES

Clevenger, Pereira, Wiechmann, Schmitt, & Schmidt Harvey (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology,* 86(3), 410-417.

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

Crane, P., Robbins, R., & Bennett, W. (2000). Using distributed mission training to augment flight lead upgrade training. In, Proceedings of Interservice/Industry Training Systems and Education Conference (I/ITSEC). (pp. 1175-1184). Orlando, FL: National Security Industrial Association.

Endsley, M. A., Selcon, S. J., Hardiman, T. D., Croft, D. G. (1998). A Comparative Analysis of SAGAT and SART for Evaluations of Situation Awareness. Paper presented at *42nd Annual Meeting of the Human Factors and Ergonomics Society,* Chicago, IL.

Endsley, M. A., Sollenberger, R. & Stein, E. (2000). Situation Awareness: A Comparison of Measures. In Proceedings of the Human Performance, Situation Awareness and Automation: User Centered Design for the New Millennium Conference, (October 2000).

Gehr, S.E., Schreiber, B.S., & Bennett, W., Jr. (2004). Within-Simulator Training Effectiveness Evaluation. Paper presented at *Interservice/Industry Training, Simulation and Education Conference (I/ITSEC 2004), Orlando, FL*

Hanson, M. A. & Borman, W. C. (1993*). Development and construct validation of the situational judgment test (SJT).* Minneapolis, MN: Personnel Decisions Research Institutes, Inc.

Hedge, J. W., Hanson, M. A., Borman, W.C., Bruskiewicz, K. T., & Logan, K., K. (1996). *Predicting the crew resource management skills of Air Force pilots* (Institute Report # 283). Tampa, FL: Personnel Decisions Research Institutes, Inc.

Krusmark, M., Schreiber, B. T., & Bennett, W. Jr. (2004). The Effectiveness of a Traditional Gradesheet for Measuring Air Combat Team Performance in Simulated Distributed Mission Operations. (AFRL-HE-AZ-TR-2004-0090). Air Force Research Laboratory, AZ: Warfighter Readiness Research Division.

Paullin, C., McKee, A., F. Hanson, M. A., & Hedge, J. W. (1994). *Development of criterion measures to assess retention and decay of aerospace physiology knowledge and skills* (Technical Report, # 255). Minneapolis, MN: Personnel Decisions Research Institutes, Inc.

Paullin, C., McKee, A., F. Houston, J., S., Hanson, M. A., & Hedge, J. W. (1997). *Development of criterion measures to assess retention and decay of aerospace physiology knowledge and skills.* (Technical Report, # 298). Minneapolis, MN: Personnel Decisions Research Institutes, Inc.

Schreiber, B.T., Watz, E.A, & Bennett, W., Jr. (2003).Objective Human Performance Measurement in a Distributed Environment: Tomorrow's Needs. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC 2003).

Schreiber, B.T., Watz, E.A, Bennett, W., Jr., & Portrey, A. (2003). Development of a Distributed Mission Training Automated Performance Tracking System. In Proceedings of the 12th Conference on Behavior Representation in Modeling and Simulation. Scottsdale, AZ.