# AN INNOVATIVE APPROACH FOR ASSESSING KNOWLEDGE IN AIR-TO-AIR DISTRIBUTED MISSION TRAINING

**Anna Castillo, Winston Bennett, Jr., Brenda Wenzel, Michael Park**
**Air Force Research Laboratory Human Effectiveness Directorate**
**Warfighter Training Research Division (AFRL/HEA)**
**Mesa, Arizona**

**Roger Schvaneveldt**
**Arizona State University**
**Mesa, Arizona**

**Robert Robbins, Jack Wooster**
**Link Simulation and Training Division**
**L3 Communications**
**Air Force Research Laboratory**
**Mesa, Arizona**

**Sarah Kotte**
**United States Air Force Academy**
**Colorado Springs, Colorado**

**ABSTRACT**

The Air Force Research Laboratory, Human Effectiveness Directorate is using its high fidelity distributed mission training (DMT) simulation testbed to explore the impact of principled training on individual and team performance. One area of interest is the development of methods for assessing the impact of distributed mission training on pilots knowledge and understanding. In previous studies we have used traditional knowledge assessment methods, which have included paper-based fill-in-the-blank tests and computer-based concept rating tasks, pre- and post-training. With the development and definition of Mission Essential Competencies (MECs) as a novel way to define complex air combat mission proficiency, these more traditional approaches to knowledge assessment and learning are not at a level of specificity for measurement and proficiency diagnosis. This paper highlights the development and lessons learned from a vignette-based approach to knowledge assessment. Our initial development which is based on Situational Judgment Inventory (SJI) and Job Knowledge Inventory (JKI) research, used an open-ended paper-based assessment instrument, referred to as Situation Assessment and Action Selection (SAAS), to examine pilots assessment of air-to-air situations as well as their opinions on appropriate courses of action. Scoring of pilot responses was challenging. One limiting factor in using open-ended responses is the time and effort required to score them. We are exploring the use of automated scoring of the responses, beginning with Latent-Semantic Analysis (LSA). Successful LSA scoring would greatly enhance the utility of the method and support the next phase of development. The next phase of development is intended to be a more automated version of the instrument, referred to as the Air Superiority Knowledge Assessment System (ASKAS). Results from our evaluation of SAAS are presented and discussed. Lessons learned and a rationale for developing a multimedia-based assessment system is discussed. Finally, key features of ASKAS are described with respect to their potential for helping researchers and practitioners assess the impact of DMT on pilots knowledge and understanding.

## Biographies

Anna Castillo is a Research Psychologist at The Air Force Research Laboratory, Warfighter Training Research Division in Mesa, Arizona. She currently serves as a principal investigator for research and development in instructional methods for aircrew training. Her research goals and objectives are to increase Distributed Mission Training (DMT) effectiveness, utilize a DMT environment to enhance learning and performance, and determine the extent to which lessons learned in DMT transfer to the operational unit. Anna received a Master

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*



| | | |
|---|---|---|
| 1. REPORT DATE<br>**DEC 2002** | 2. REPORT TYPE<br>**Conference Proceedings** | 3. DATES COVERED<br>**01-01-2001 to 31-10-2002** |

| | |
|---|---|
| 4. TITLE AND SUBTITLE<br>**An Innovative Approach for Assessing Knowledge in Air-to-Air Distributed Mission Training** | 5a. CONTRACT NUMBER<br>**F41625-97-D-5000** |
| | 5b. GRANT NUMBER |
| | 5c. PROGRAM ELEMENT NUMBER<br>**62205F** |
| 6. AUTHOR(S)<br>**Anna Castillo; Winston Bennett, Jr.; Brenda Wenzel; Michael Park; Roger Schvaneveldt** | 5d. PROJECT NUMBER<br>**1123** |
| | 5e. TASK NUMBER<br>**B0** |
| | 5f. WORK UNIT NUMBER<br>**01** |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>**Air Force Research Laboratory/RHA,Warfighter Readiness Research Division,6030 South Kent Street,Mesa,AZ,85212-6061** | 8. PERFORMING ORGANIZATION REPORT NUMBER<br>**AFRL; AFRL/RHA** |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)<br>**Air Force Research Laboratory/RHA, Warfighter Readiness Research Division, 6030 South Kent Street, Mesa, AZ, 85212-6061** | 10. SPONSOR/MONITOR'S ACRONYM(S)<br>**AFRL; AFRL/RHA;** |
| | 11. SPONSOR/MONITOR'S REPORT NUMBER(S)<br>**AFRL-RH-AZ-PR-2002-0010** |

12. DISTRIBUTION/AVAILABILITY STATEMENT
**Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES
**In Proceedings of the 2002 Interservice/Industry Training, Simulation, and Education Conference, held in Orlando FL on 2-5 Dec 02 *Additional authors: Robbins, Robert; Wooster, Jack, & Kotte, Sarah**

14. ABSTRACT

**This paper highlights the development and lessons learned from a vignette-based approach to knowledge assessment. Our initial development which is based on Situational Judgment Inventory (SJI) and Job Knowledge Inventory (JKI) research, used an open-ended paper-based assessment instrument, referred to as Situation Assessment and Action Selection (SAAS), to examine pilots assessment of air-to-air situations as well as their opinions on appropriate courses of action. Scoring of pilot responses was challenging. One limiting factor in using open-ended responses is the time and effort required to score them. We are exploring the use of automated scoring of the responses, beginning with Latent-Semantic Analysis (LSA). Successful LSA scoring would greatly enhance the utility of the method and support the next phase of development. The next phase of development is intended to be a more automated version of the instrument, referred to as the Air Superiority Knowledge Assessment System (ASKAS). Results from our evaluation of SAAS are presented and discussed. Lessons learned and a rationale for developing a multimedia-based assessment system is discussed. Finally, key features of ASKAS are described with respect to their potential for helping researchers and practitioners assess the impact of DMT on pilots knowledge and understanding.**

15. SUBJECT TERMS

**Knowledge assessment; Air-to-air distributed mission training; Distributed mission training; Flight simulation; Team performance; Individual performance; Latent Semantic Analysis (LSA); Air Superiority Knowledge Assessment System (ASKAS); Learning and performance;**

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | |
| **unclassified** | **unclassified** | **unclassified** | **Public Release** | **8** | |

of Science in Experimental Psychology from New Mexico Highlands University in 1998 and a Bachelor of Arts in Journalism / Mass Communications from New Mexico State University in 1995.

Dr Winston Bennett, Jr. is a Senior Research Psychologist with the Air Force Research Laboratory Human Effectiveness Directorate, Warfighter Training Research Division, located at Williams-Gateway Airport, Mesa AZ. He is the team leader for training systems technology and performance assessment research and development. He received his Ph.D. in Industrial Organizational Psychology from Texas A&M University in 1995. Dr. Bennett has published numerous research articles, book chapters and technical reports in the Human Resources arena. He is actively involved in research related to performance evaluation, personnel assessment, training requirements identification, and quantifying the impact of organizational interventions - such as interactive, high fidelity immersive simulation environments and job redesign/restructuring and training systems - on individual, team and organizational effectiveness.

Dr. Brenda M. Wenzel is a Research Psychologist at the Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Training Research Division in Mesa, AZ. Her work spans the areas of interactive individualized training, knowledge elicitation and representation, and human/computer interaction. She earned a B.S. in Psychology from Ball State University, Muncie, Indiana. She earned both her M.A. and Ph.D. in Psychology from New Mexico State University, Las Cruces, New Mexico.

Lt Col Mike Odie Park is the Training Research Operations Officer at The Air Force Research Laboratory, Warfighter Training Research Division in Mesa, Arizona and currently is an F-16C Flying Training Unit Instructor Pilot at Luke AFB. His goal is to ensure that warfighters get the best available training to enhance their skills for combat. Lt Col Park graduated from the University of Washington in 1984 with a Bachelor of Science degree in Aeronautical and Astronautical Engineering, and a Master of Science degree in Human Resources from Troy State in 1995. He is a command pilot with over 3200 hours in F-16, A-10, and AT-38 aircraft.

Roger Schvaneveldt is a Professor of Applied Psychology at Arizona State University, East in Mesa, AZ. He earned the Ph.D. in psychology at the University of Wisconsin in 1967. His research interests include basic issues in memory and cognition as well as applied research in human factors, in general, and aviation, in particular.

Robert Robbins works as an Instructional Systems Designer for Link Simulation and Training at the Air Force Research Laboratory in Mesa, AZ. He is a retired Air Force fighter pilot with over 3800 total flight hours, over half as a flight examiner/instructor pilot. He has a Masters degree in Aeronautical Science.

Jack Wooster works as a subject matter expert for Link Simulation and Training at the Air Force Research Laboratory in Mesa, AZ. He is a retired Air Force fighter pilot with over 3500 hours in fighter aircraft and 2500 hours in the F-16. He has a Bachelor of Science in Aerospace Engineering from the University of Michigan and a Masters in Aeronautical Science from Embry-Riddle Aeronautical University.

Sarah Kotte is a Cadet at the United States Air Force Academy. She is pursuing her Bachelor of Science in Behavioral Sciences. Her research interests include cognition and applied research in human factors.

# AN INNOVATIVE APPROACH FOR ASSESSING KNOWLEDGE IN AIR-TO-AIR DISTRIBUTED MISSION TRAINING

**Anna R. Castillo, Winston Bennett, Jr., Brenda Wenzel, Michael Park**
**Air Force Research Laboratory Human Effectiveness Directorate**
**Warfighter Training Research Division (AFRL/HEA)**
**Mesa, Arizona**

**Roger Schvaneveldt**
**Arizona State University**
**Mesa, Arizona**

**Robert Robbins, Jack Wooster**
**Link Simulation and Training Division**
**L3 Communications**
**Air Force Research Laboratory**
**Mesa, Arizona**

**Sarah Kotte**
**United States Air Force Academy**
**Colorado Springs, Colorado**

## INTRODUCTION

The Air Force Research Laboratory (AFRL), Warfighter Training Research Division has been participating in air-to-air Distributed Mission Training (DMT) research and development efforts with AFRL s networked 4-ship F-16 testbed in Mesa, Arizona since 1997. Over the years, various methods of data collection and assessment methodologies have been utilized. Previous research has demonstrated that DMT can provide effective training tailored to meet defined learning objectives through careful development and delivery of scenarios that are presented in a building block format over several training sessions (Bennett & Crane, 2002).

In an attempt to evaluate the impact of training, researchers examined changes in knowledge as a function of training. A goal of DMT is to produce expertise in performance in flight. Expert performance depends on the acquisition of both knowledge and skill (Schvaneveldt, Tucker, Castillo, & Bennett, 2001). In earlier work, we have examined knowledge change using indirect methods employing networks of pilot knowledge (Schvaneveldt, Tucker, Castillo, & Bennett, 2001). That investigation showed that less experienced pilots demonstrated reliable changes in the way they organize concepts pertaining to air-to-air combat missions after a week of training in DMT s high-fidelity simulators. Their knowledge networks were more like the networks of experienced pilots at the end of the week compared to the beginning of the week. It is also valuable to pursue the study of knowledge change using more direct methods of assessing pilots understanding of particular aspects of air-to-air combat scenarios.

While these network assessments provide useful criterion data on the impact of training on overall learning, they do not permit detailed assessments of particular competencies, knowledge and skills that underlie the observed changes in networks over the course of a week of training or after some transfer interval to the field. What s needed is an innovative and robust assessment system that can link performance to proficiencies on critical knowledge, skills, experiences and competencies associated with complex combat missions.

This paper describes a method of knowledge assessment referred to as Situation Assessment and Action Selection (SAAS). The approach used to develop SAAS comes from research on the development and validation of Situational Judgment Inventories (SJIs) and Job Knowledge Inventories (JKIs) (see Hanson & Borman, 1993; Hanson, & Hedge, 1994; 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 proven 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 situationally-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). SAAS represents a first attempt to assess the feasibility of using SJI and JKI-like paper-based assessment methods to quantify specific learning benefits in a complex air combat domain.

## DEVELOPMENT AND EVALUATION OF SAAS

Our SAAS instrument was developed by researchers and subjects matter experts (SMEs) at AFRL Mesa and was the outcome of a series of workshops and discussions regarding the type of knowledge gained in a DMT environment and how best to assess this knowledge. SAAS was designed to (a) determine participants baseline knowledge of the subject matter prior to engaging in DMT; (b) motivate the participant to acquire new knowledge; (c) help determine the extent to which progress has been made in achieving the training objectives; and (d) measure new knowledge gained by the end of a week of nine structured sorties in the DMT testbed environment. Results from the analysis of SAAS have contributed to specifications for the next generation of knowledge assessment research.

Our use of more traditional approaches to measuring learning and performance has provided us with extremely useful data regarding the over all learning that can occur as a result of principled strategies and syllabi in DMT. With the advent of Mission Essential Competency (MEC) development research with Air Combat Comment (ACC), a greater level of measurement specificity is required in order to track proficiency at the finer grained analysis afforded by the specification of MECs. A MEC is the knowledge, skill, ability, or experience that is necessary to achieve successful performance in a given mission element (Bennett, Schreiber & Andrews, in press; Colegrove & Alliger, 2002). The identification of these skills is critical in that it allows researchers to focus mission training objectives on very specific aspects of competency development and to potentially measure the extent to which the training system can aid in developing targeted skills in training and in operational transfer environments.

## DMT EXERCISES

DMT research exercises typically last for four and one-half days allowing teams to fly nine, one-hour missions or sorties . Pilots participating in DMT fly two missions per day on Monday, Tuesday, Wednesday, and Thursday, and fly one morning mission on Friday. This schedule supports a building-block (crawl — walk — run) approach to training in which learning objectives for missions later during the week are dependent upon mastery of skills exercised earlier (Bennett & Crane, 2002). Three DMT syllabi have been designed to expose the participants to scenarios of increasing levels of complexity. Research protocol consists of standardized benchmarks on Monday afternoon and Friday morning. Benchmarks are defensive counter air (DCA) point defense missions (same mission type as the SAAS scenarios). Monday s benchmarks are extremely difficult for all groups, however by Friday the learning curve is such that their overall performance is noticeably higher. Both the number and intensity of the threats surpass what the participants have previously been exposed to in normal flying training. The notable improvements on Friday s vs. Monday s benchmarks demonstrates the manner in which this training strategy is conducive to enhanced air-to-air awareness and subsequent improvement in mission performance (Bennett, et al., 2002).

Interviews with SMEs who observe and evaluate mission performance in the testbed were asked about the benefits of DMT as a training research tool. They noted that as a result of DMT exercises, participants are better able to listen, assess information, and execute their briefed communication and tactical gameplan.

Important benefits of concentrated air-to-air training in this capacity include focus on briefing, execution, debriefing, and correcting execution errors through lessons learned in debrief. Participants have the opportunity to improve on / implement what they learned from debrief on subsequent missions. Intense repetition of 4 V 4 and 4 V X engagements is rarely (if ever) practiced operationally due to resource and airspace constraints in primary training. Tactic shifts may be based on that knowledge rather than contingencies.

Participants complete an after action survey that gives them an opportunity to articulate strengths and weaknesses of the system, benefits gained, lessons learned, etc. When asked what they have gained from participating in DMT, some of the most mentioned skills include:

Validation of tactics
Confidence in decision-making
Improvement in overall SA
Better shot discipline
Better awareness of AWACS / WD limitations
Appreciative of pace of missions and progression of complexity

Through the data obtained from SAAS, we hope to quantify this noticeable increase in mission performance by identifying specific skills that are enhanced through immersion in the simulation system.

## SAAS ADMINISTRATION METHOD AND SCENARIOS

Participating pilots reported F-16 flying hours from 80 to 2600. Participants completed the SAAS pre- and post-DMT. Parallel forms of SAAS (versions A and B) were created to control for potential practice effects associated with test-rest. The forms were counterbalanced across participants with each participant completing both versions.

The SAAS instructions and scenarios are as follows: *In this exercise, we would like you to tell us how you would approach a particular air-to-air combat situation by writing a summary of your tactics and game plan. On the following page is a depiction of a situation showing the positions of bogey and/or hostile aircraft in the airspace relative to your Viper 4-ship. Assume you are on a Defensive Counter Air (DCA) Point Defense Mission defending your airfield. Adversary airspeed is between 350C and 1.2 mach. You load out is 4 X 2 X gun with 2 wing tanks. Your initial speed is 350C.*

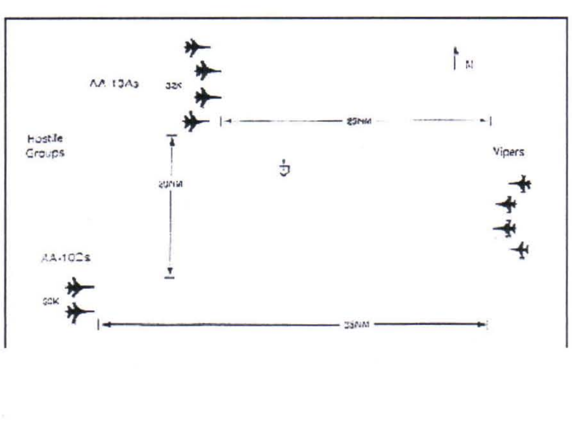


Figure 1. Version A of the SAAS depicting a 4V6 DCA point defense mission. The scenario consists of a two-group Azimuth presentation. Both groups are initially positioned west of bullseye. The North group is heavy and consists of four SU-27s in a line abreast formation carrying AA-10 Alpha missiles. The South group is echelon SW from North group and consists of two SU-27s in a line abreast formation armed with AA-10 Charlie missiles.

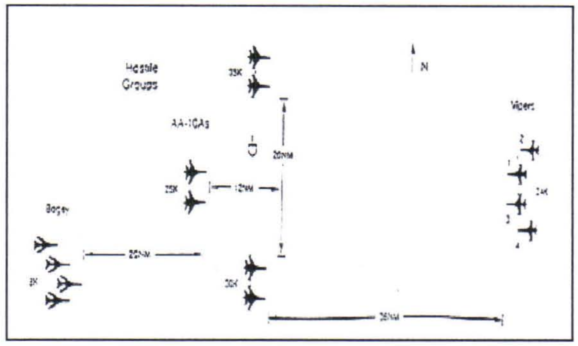


Figure 2. Version B of the SAAS presents a two-package picture consisting of four groups. The lead package is a three-group Champaign consisting of SU-27s armed with AA-10 Alpha missiles. The lead groups are adjacent to bullseye. The second package consists of a bogey group of Mig-23 striker aircraft at low altitude. This is a 4V10 DCA point defense mission.

## SCORING SCHEME

The scheme used to score SAAS responses is presented in Figure 3. This scoring scheme was developed by the fourth author to streamline the scoring process and to develop standards against which responses could be more consistently scored. A number of difficulties were encountered in the development of this scoring scheme:

1. Both scenarios have blue fighters already at a disadvantage.
2. Adversary reaction level unknown. This has an effect on shot doctrine.
3. Although it is mentioned that blue fighters are in a DCA Point Defense role, there is no mention of length of vulnerability, previous engagements, or how long fighters have been there. This affects jettison decision and radar/missile employment.
4. There are numerous tactics that flight leads may use and there is not necessarily a correct answer. This includes not knowing acceptable level of risk.

5. The scenarios involve blue fighters starting in different positions relative to bullseye. There is no mention what the fighters are protecting and where it is located. This has implications for desired engagement zone / gameplan.

The following criteria were used to grade SAAS scenarios. Points are not necessarily relative to the importance of the question, but were assigned based on the available choices to be made. **13 total points available.**

1. QUESTION: What action would you take at the commit? **(4 points possible)**

   Power = 1 point
   Action = 1 point
   Altitude = 1 point
   Airspeed = 1 point

2. QUESTION: How would you target this picture? **(4 points possible)**

   1 point for each number mentioned
   (even if only sanitizing)

3. QUESTION: What tactics / gameplan would you employ? **(3 points possible)**

   Mention of valid gameplan = 1 point
   Backup gameplan
   (adversary maneuvers, or fighter pause) = 1 point
   Shot criteria / support of shot = 1 point

4. QUESTION: Would you jettison your tanks? If so, when? **(2 points possible)**

   No = 0 points
   Yes = 1 point (commit, merging, adversary inside certain range during egress)

   Mention of high fast flyer = 1 additional point

Figure 3. SAAS Scoring Scheme Breakdown

After examining these difficulties, researchers next turned to examining the sensitivity of SAAS to flying experience levels and to also consider an alternative, and potentially easier, means of scoring the responses.

## RESULTS

Over 150 SAAS were scored using this method and a brief description of the results follow. Due to incomplete data (both pre- and post-versions not completed, early departures, etc.) there were only 130 valid cases (65 pre- and 65 post-) included in the analysis.

An analysis of variance revealed a significant interaction of experience and pre- vs. post-test scores ($F(1, 61) = 4.269$, $p=.043$). The interaction is shown in Table 1 below. In this study, novice pilots were considered those with 500 hours in the F-16 and below. Experienced pilots were those with over 500 hours in the F-16. While novices show improved performance after a week of training, experienced pilots actually score worse at the end of the week than at the beginning. No other effects were significant. The different versions of the test were roughly equivalent. Although, there may be some differences in how novices and experienced pilots deal with the two different scenarios, we suspect that the poorer performance by the experienced pilots at the end of the week may reflect a failure of the experienced pilots to take the second test seriously. An alternative explanation is related to the principled nature of the training in our research environment and its impact on traditional approaches to weapons employment, which might have been manifest in their pre-test performance.

The Air Force currently equates mission-qualified experience to the total number of flying hours in the given weapons system —not on the content or quality of the hours. It is very conceivable that the results from the experienced pilots post-test scores might be indicative of having been exposed to a competency-based syllabus where their past live-fly experiences were challenged and potentially changed. There was a significant difference between novice and experienced pilots in the pre-test Scores ($F(1,61)=4.863$, $p=.031$) indicating that the test is sensitive to experience. Further work is presently underway to clarify the changes that occur over training.

Table 1. Mean SASS Scores as a Function of Experience and Time of Test

| | Experienced | Novice | Mean |
|---|---|---|---|
| PreTest | 9.35 | 8.31 | 8.83 |
| PostTest | 8.76 | 8.85 | 8.81 |
| Mean | 9.05 | 8.58 | 8.82 |

## ALTERNATIVE SAAS SCORING METHOD: LATENT SEMANTIC ANALYSIS (LSA)

LSA is a machine-learning method for automatically extracting and representing knowledge in massive databases of relevant electronic text (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990). It was developed through ten years of basic and applied research supported by Bell Communications Research, DARPA, ONR, ARI, NASA, AFRL, the McDonnell Foundation and others. LSA has been extensively validated in both controlled experiments and field tests

(Landauer & Dumais, 1997; Landauer, Foltz, and Laham, 1998; Landauer, 1998).

We are interested in utilizing this method to objectively compare SAAS responses and search for trends. In order to run the SAAS data through LSA, it must first be tagged in Extensible Markup Language (XML). An example set of responses can be seen in Figure 4. LSA has a variety of applications to text-based research. The ability to conduct matching at a quantifiable semantic level between pieces of text material, allows LSA to perform analyses that were formerly only done through hand-coding. Results comparing LSA s predictions with hand-coding indicate that the percent agreement between LSA and humans is close to the percent agreement between human coders (Foltz, 1996). Using LSA for the SAAS data is currently in a proof of concept phase. Successful LSA scoring would greatly enhance the utility of the method and support the next phase of development of which we hope to present results next year.


```
<saas version= * id=   date=   Unit=   grade=  >
  <risk>
      Commanders intent/what you're protecting
      Your ordinance vs adversary ordinance/observed
      tactics
      Other assets (other air, other ground based
      defenses)
      Location of engagement (in front of, in, behind)
      desired engagement zone
  </risk>
  <action score=  >
    Gate, go out, accelerate supersonic and climb to 30-
    35K. Once in a position of advantage, recommit
    back in as well (but not necessarily visual wall). See
    tank discussion below.
  </action>
  <target score=  >
    2 to north lead group, 4 to south lead group. 1 and 3
    fill in appropriately to leading edge. WD targets trail
    group and low bogey group (if he can detect).
    Assuming destruction of leading edge (north and
    south lead group) then #2 to trail group. After trail
    group dead, #2 and 4 bogey group with 1 and 3
    filling in at 20 NM.
  </target>
```


Figure 4: Example of XML Tagged SAAS Response

## LESSONS LEARNED

Results from the preliminary analysis utilizing the scoring scheme revealed the need to develop a more robust instrument that allows for fewer assumptions about the scenario itself as well as a more definitive method of scoring. As well as providing a springboard for more innovative methods of scoring, such as LSA, preliminary SAAS data were very useful as they inspired brainstorming on a diagnostic, multimedia, automated version of the instrument. If the goal of the SAAS is to identify the skills that are enhanced in a dynamic and immersive learning environment such as DMT, then we should be able to identify potential skill deficiencies, provide this feedback to the participants, and provide them tools with which to target these skills throughout the course of the week.

The goal of examining pilots assessment of air-to-air situations as well as their opinions on appropriate courses of action was realized through SAAS. Lessons learned from this exercise will have a significant impact on future assessment methodology and research protocol when attempting to study knowledge acquisition. The unexpected pattern of pre- and post-scores indicates a need to move to a more sensitive measure of pilot knowledge. In addition, it was recognized by both participants and evaluators that the scenarios did not provide enough information to reduce the number of assumptions that need to be made in order to make an accurate assessment of the situation.

Given the results and lessons learned from our SAAS evaluation, it is obvious that the dynamic and complex nature of the domain dictates a more robust approach to the level of specificity in assessment that must be achieved. This idea lead to specification development of what we are currently calling the Air Superiority Knowledge Assessment System (ASKAS). ASKAS represents a further extension of both the SJI and JKI research methodologies and uses automation for situation or scenario item, and knowledge item, presentation and for response elicitation and tracking. Eventually, it will also include an online scoring capability.

With our approach to ASKAS, we will link a competency-based air combat SJI and JKI to specific learning objectives. We will then be able to efficiently assess a variety of combat-relevant knowledge, skills and competencies and to demonstrate an extremely high fidelity assessment capability that does not exist today.

The ASKAS project is in initial design / development phase. ASKAS is a logical extension to our SAS research and our attempts to address some of the more salient difficulties we encountered with the paper-based assessment. The ASKAS research effort will involve using computer-based multi-media vignettes of specific DMT scenarios. The goal in using a more robust multi-media approach to the assessment is proving the pilot with a more complete representation of the flow and crucial triggers and events of the particular scenario. The feedback from pilots using SAAS indicated that the static, snapshot representation of the scenario did not provide enough of a context for them to appropriately

Table 2. A Comparison Of The First And Next Generation Assessment Methodologies Highlight The Pros And Cons Of Paper-Based Vs. Computer-Based Assessments

| **Situation Assessment and Action Selection (SAAS)** |
|---|
| Gauges pilot s existing air combat knowledge |
| Used to assess situational knowledge gained in DMT |
| Time lag between administration and scoring |
| No feedback to pilots |
| Subjective scoring |
| Ambiguities / assumptions in scenario impact scoring |
| **Air Combat Situation Knowledge Assessment System (ASKAS)** |
| Assess pilot knowledge and understanding of critical situations and mission features based on MECs |
| Multimedia platform; Web administration capability |
| Scored in real-time |
| Provides immediate feedback |
| Deployable to the field |
| Diagnostic capabilities |

respond. The multi-media approach permits us to examine the entirety of a scenario and to obtain assessments at various stages of the scenario as it unfolds and to focus the assessment on different competencies, knowledge and skill proficiency as a function of its relevance for that particular portion of the scenario.

Scenarios being considered for ASKAS would be representations of actual real time missions captured to a file complete with radio communication. Questions would consist of multiple choice questions or short answers related to specific aspects of the scenario at a given time in the flow of the scenario.

The automation of the ASKAS process also permits us to systematically link the response to the questions we ask, to a very specific portion of the scenario where it will be possible for expert scorers to identify the most and least appropriate responses to the scenario at that point in time. This type of systematic and controlled linkage of events to criteria simply isn t possible with a static, paper-based form of the scenario. Moreover, we feel it will be possible to identify expert scoring schemes, which we can then automate in the ASKAS software to facilitate more responsive assessment and diagnosis. Table 2 presents a comparative assessment of the benefits of the proposed new measure, ASKAS.

## CONCLUSIONS AND NEXT STEPS

It is clear that the complexity of the air combat domain does not lend itself to the more straightforward assessment approach afforded us with SAAS. Moreover, this domain complexity indicates that a more robust and context-driven approach, such as that proposed with ASKAS, may be the only reliable and valid way to achieve the level of measurement precision we need for future DMT training diagnosis and assessment. When the multi-media version of the instrument is implemented, researchers may wish to administer the post-test Friday morning prior to the last mission of the week.

Further research needs to be conducted to determine the degree to which (if any) giving the post test one mission early affects assessment outcome. Another issue that needs to be overcome in ASKAS is the fact that some answers are not mutually exclusive. Subject matter experts indicate that there can be more than one right answer and techniques among fighter pilots tend to vary depending on where and when they were trained. Also, the assumptions that had to be made to complete the instrument may have had an effect on the demonstration of variability from the beginning of the week to the end. It is evident that lessons learned in SAAS will be conducive to a more stringent assessment tool. Therefore, future DMT participants who come to AFRL can look forward to participating in some cutting-edge state- of -the -art training research that will help enhance their skills both in simulated as well as live fly.

## REFERENCES


Bennett, W. Jr. & Crane, P. (2002). The deliberate application of principles of learning and training strategies within DMT. Paper presented at the NATO SAS-038 Working Group Meeting, Brussels, Belgium.

Bennett, W. Jr., Schreiber, B. T., & Andrews, D. H. (in-press). Developing competency-based methods for near-real-time assessment in a distributed training environment. Computers in Human Behavior.

Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Schmidt Harvey, V. (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 mission requirements in a novel way. Paper presented at the NATO SAS-038 Working Group Meeting, Brussels, Belgium.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing By Latent Semantic Analysis. Journal of the American Society For Information Science, 41(6), 391-407.

Foltz, P. W. (1996). Latent Semantic Analysis for text-based research. Behavior Research Methods, Instruments and Computers. 28(2), 197-202.

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.

Landauer, T. K. (1998). Learning and representing verbal meaning: The Latent Semantic Analysis theory. Current Directions in Psychological Science, 7, 161-164.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. Psychological Review, 104, 211-240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284.

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.

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.

Rowe, A. L., Cooke, N. J., Hall, E. P. & Halgren, T. L. (1996). Toward an online assessment methodology: Building on the relationship between knowing and doing. Journal of Experimental Psychology: Applied, 2, 31-47.

Schvaneveldt, R., Tucker, R., Bennett, W., Jr., & Castillo, A. (2001). Knowledge acquisition in distributed mission training. Paper presented at the annual I/ITSEC Conference, Orlando, FL.

Schvaneveldt, R., Durso, F., Goldsmith, T., Breen, T., Cooke, N., Tucker, R., & DeMaio, J. (1985). Measuring the structure of expertise. International Journal of Man-Machine Studies, 23, 699-728.

Tessmer, M., Perrin, B., & Bennett, W., Jr. (1997). Assessing the stability of structural learning measures (AL/HR-TR-1997-0131). Mesa AZ: Air Force Research Laboratory, Warfighter Training Research Division.