THE INTERNATIONAL JOURNAL OF AVIATION PSYCHOLOGY, *4*(3), 241–264

Individual Differences in Pilot Situation Awareness

Mica R. Endsley
Texas Tech University, Lubbock

Cheryl A. Bolstad
Monterey Technologies, Inc.
Cary, North Carolina

Although anecdotal evidence has suggested considerable individual differences in the abilities of pilots to acquire and maintain situation awareness (SA), specific research to validate this claim and investigate the locus of such differences is lacking. This article presents an initial investigation of individual differences in SA. A study was conducted in which experienced fighter pilots completed a battery of tests to measure their abilities along key dimensions hypothesized to be important for SA. These measures were compared to subjects' abilities in situation awareness. The presence of consistent individual differences in SA abilities was supported and several key abilities were identified.

The objective of this study was to determine whether situation awareness (SA) abilities vary in any reliably consistent manner between individuals and to identify explicitly those characteristics that may contribute to high SA in individuals. A tremendous amount of emphasis has been placed on the necessity for aircrew members to have high SA. *SA,* a person's mental model of the world, can be defined formally as "the perception of the elements in the environment, within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (Endsley, 1988b).

Although the recognized need for SA can be traced back to World War II (Press, 1986), an increased focus on the construct has arisen in recent years,

Requests for reprints should be sent to Mica R. Endsley, Department of Industrial Enginering, Texas Tech University, Lubbock, TX 79409–3061.

largely due to the vastly increased complexity of aircraft systems and the flight environment. Although most efforts have sought to improve SA through the aircraft design process, it may also be possible to improve SA at the level of the individual through training or selection. This would allow improvements in aircrew SA to be made with existing systems and personnel, as well as in new aircraft being developed.

To make any improvements along these lines, however, it is first necessary to determine which factors allow one person to achieve better SA than another from the same set of information and displays. A considerable amount of variance between individual pilots has been demonstrated at the performance level (Youngling, Levine, Mocharnuk, & Weston, 1976). At least part of this difference may be attributable to individual differences in SA capabilities; evidence for this, however, has remained anecdotal at best.

Figure 1 provides a model of the role SA plays in the decision process. As shown in Figure 1, in addition to external task and environmental factors, SA is believed to be affected by a variety of individual characteristics. Among these are experience, training, and individual attributes. Experience should be instrumental in improving an individual's SA through a variety of mechanisms. First, much of SA may derive from pattern-matching environmental features to structures in long-term memory, thus providing the higher levels of SA (comprehension and projection) and links to decision information. This would occur first through the development of a larger body of episodic memories to draw upon for pattern-matching and solution selection. Increased experience in the environment should lead to the formation of higher level structures such as schemata or mental models, which can be used to organize the complexity and multiplicity of objects in the environment. These schemata and mental models become richer with further experience to include relations between objects, functioning, and cause and effect information, to the point of allowing accurate projection of future states (the highest level of SA) and the development of appropriate response scripts linked to the schemata. Another effect of experience is that important environmental cues that signify relevant information can be identified (consciously or unconsciously), providing better information and keying the appropriate schema in memory.

Experience may also improve SA by reducing the amount of operator resources required for specific tasks, thus freeing up resources for achieving SA. Experience should lead specifically to a decreased requirement for attention to individual tasks and an increase in attention-sharing between tasks. Experience is also necessary for the development of automaticity for tasks, which further reduces task attention requirements.

Training can be a useful mechanism for assisting in developing pilots with superior SA abilities by structuring and focusing the experience process. Many new ideas for directly improving SA through enhanced training programs have been suggested (Endsley, 1989b; Kass, Herschler, & Companion, 1990). To achieve real progress in this area, however, it is first necessary to

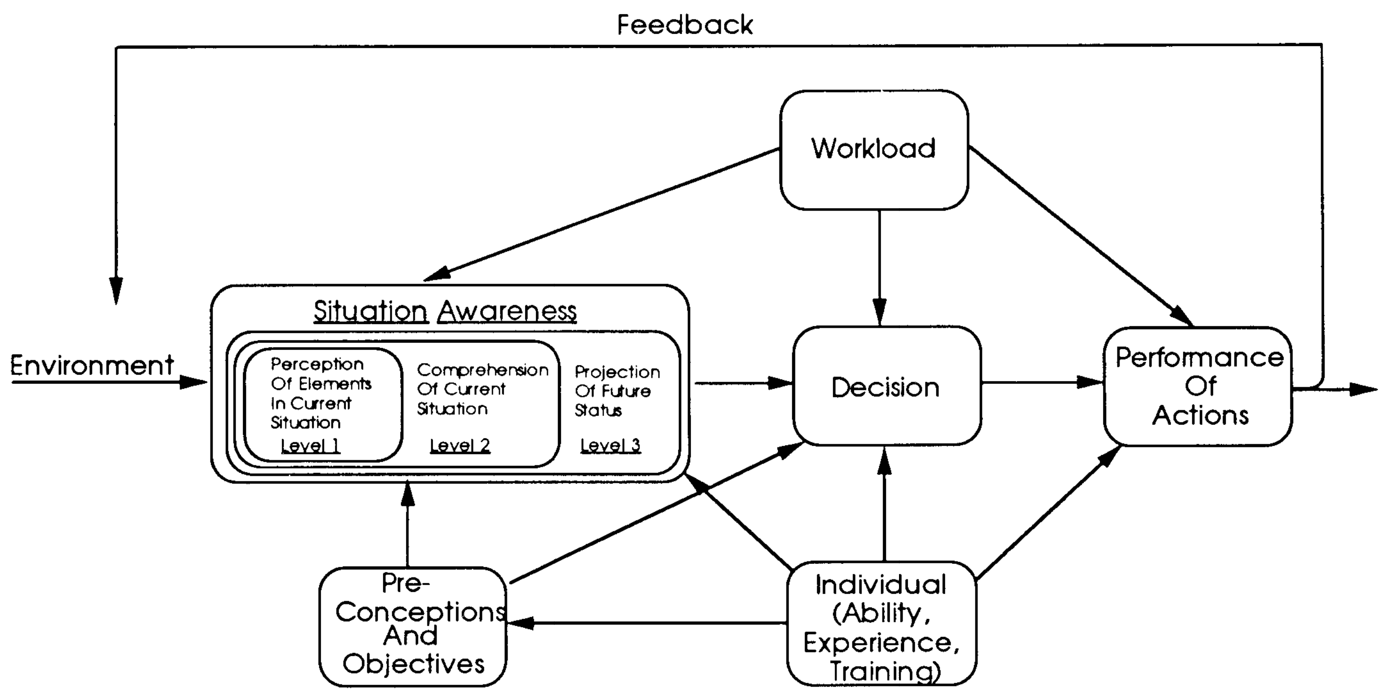

FIGURE 1 Role of situation awareness in the decision process adapted from Endsley (1988b).

identify the specific skills that need to be trained. That objective may be at least partially fulfilled by identifying those skills or attributes that separate really successful pilots from the rest of the crowd. Endsley (1988a) identified five primary areas that may relate to individual differences in SA: spatial, attention, memory, perception, and cognitive functions.

AREAS OF INDIVIDUAL DIFFERENCE

Spatial

The degree to which an individual can mentally visualize and manipulate objects spatially and visualize one's own orientation relative to those objects is probably important to SA in aircraft systems. Much of SA—particularly in the flight environment—relies on understanding spatial relationships between dynamic objects in three-dimensional space. Spatial abilities are also needed for navigational purposes. The ability to hold a mental map of the environment may be fundamental to some components of SA. Thorndyke and Stasz (1980) have found substantial individual differences in map learning, which were correlated with visual-spatial abilities and field-independence. Many pilot-selection batteries employ tests of spatial ability, including Lufthansa of Germany (Kirsch, 1976) and the U.S. Air Force Basic Attributes Test (BAT; Carretta, 1987a; Kantor & Bordelow, 1985), due to its importance in the flight environment. For instance, Carretta (1987b) found that performance on a mental rotation test was related to recommendations for fast-jet training after undergraduate pilot training.

Attention

Capabilities involving attention and attention sharing are important to achieving good SA in a demanding environment. The distribution of attention across multiple, competing sources of information and tasks is a significant challenge for aircrews. Several authors have found correlations between time-sharing capabilities and flight performance (Damos, 1978; North & Gopher, 1976). Additionally, Gopher and Kahneman (1971) found that selective-attention abilities were correlated with pilot success in the Israeli Air Force.

Memory

SA may well be related to working memory capacity and the quantity and quality of long-term memory stores. Working memory is called upon a great deal in the SA process. In the absence of other mechanisms, the higher levels of SA—comprehension and projection of future events—must occur in working memory as people try to integrate information from multiple sources, compare this information to goals and objectives, and project future scenarios from known dynamics. Working memory is further taxed by simultaneous decision making and carrying out other flight tasks. Long-term memory stores, where they exist, may substantially reduce this load on working memory.

Perception

Abilities in perception may well relate to SA, as SA relies first upon accurately perceiving necessary information. In a taxing flight environment, individuals with superior perceptual abilities may have a slight edge over others. Perceptual speed (the ability to quickly perceive information), encoding speed (the ability to quickly encode perceived information), and vigilance (the ability to remain alert to the perception of infrequent signals) may all be related to SA. Each of these skills is either part of or under consideration for inclusion in the BAT (T. Carretta, personal communication, January 25, 1989). In addition, pattern-matching skills may be frequently called upon.

Cognitive Functions

Higher order cognitive functions should be useful in helping to search out information and piece it together to form the higher levels of SA—comprehension and projection. Unless this information can be derived from existing memory stores, such analysis will probably be necessary. In addition, other factors such as cognitive complexity, field independence, and locus of control may relate to SA, because they have been linked to problem solving and workload management in a number of studies.

Cognitive complexity. Cognitive complexity ranks people into four groups, ranging from concrete to abstract. *Concrete* individuals respond to the world in a very externally oriented fashion, are very close-minded regarding their beliefs and identify highly with social roles and status positions. *Abstract* individuals are very abstract in their thinking; they tend to react to their environment based upon their internal values and beliefs. Cognitive complexity has been found to be correlated with an individual's ability to handle mental workload in more generalized problem solving (Robertson, 1984).

Field independence. Field independence refers to a person's ability to restructure problems cognitively, to perceive objects distinct from their context, or to provide a new organization to an unstructured field, as opposed to field-dependent people, who are more constrained by the prevailing organization of the problem representation (Witkin & Goodenough, 1977). Field-independent individuals have been shown to be more effective at decision making in a variety of situations (Stasz & Thorndyke, 1980; Witkin, Goodenough, & Oltman, 1977; Witkin, Moore, Goodenough, & Cox, 1977), including perspective tasks, transformation tasks, incomplete figures tasks, and map learning.

Locus of control. Additionally, locus of control may be an important factor in how likely a person is to take an active role in seeking out SA. An internally controlled person believes that a successful outcome is due to one's own behavior. An externally controlled person believes that a successful outcome is due to luck, chance, fate, or another person's actions (Rotter, 1966).

In this study, we wanted to ascertain which, if any, of these factors do indeed relate to individual abilities in SA. A test battery was therefore constructed to measure these attributes. This battery was administered to a group of experienced fighter pilots and compared to their SA as assessed during simulations of air-to-air fighter missions.

ATTRIBUTE TEST BATTERY

For the five areas identified, available tests in the literature that purport to measure these attributes were reviewed. Because the armed services have developed many applicable attribute tests for pilot selection, these test batteries were reviewed thoroughly, along with the many generally available psychometric tests. The Air Force uses the BAT, sometimes known as the PortaBAT (Carretta, 1987a). Currently, the Navy has two selection tests. The first, the Aviation Qualification Test (AQT), is given to all flight candidates. The second, the Flight Aptitude Rating (FAR), is given to potential naval pilots (Griffin, 1988). In addition, the JAMJET (Jensen, Adrion, Maresh, & Weinert, 1987), which was created for pilot evaluation, was reviewed. From our literature-search results, the military-testing literature, and discussions with professionals directly involved in military testing, hundreds of tests were found that could possibly measure SA correlates.

From this pool, selection of specific tests to include in the battery was based on several criteria. Preference was given to standardized tests and tests that are in current military-selection batteries because information on reliability and validity is readily available. Tests that were easy to develop and administer or that can be given in a computerized format were also given a high priority for selection. Of over 100 tests reviewed, 18 were selected for inclusion in the battery with at least one test in each of the five attribute areas, as listed in Table 1. Although there were many tests that may be construed as potentially applicable, these were selected for this initial examination of attributes.

Many of the tests included in the SA attribute battery were similar to the tests used in the current BAT or JAMJET. The AQT and the FAR were not used, because the Navy has made few changes to the battery since 1947 and a new, automated naval battery is expected in the near future. Several additional tests, consisting mostly of modifications of standardized tests or psychological paradigms, were included in the battery. The final battery used two methods for administration: paper and pencil and computer. Details of the SA attribute battery can be found in Bolstad and Rodriquez (1991) and Bolstad (1991). Summaries of the selected tests follow.

Spatial Tests

Four tests were selected to measure an individual's ability to mentally visualize and manipulate objects spatially. Each was presented as a paper-and-pencil task.

TABLE 1
Situation Awareness Attribute Test Battery

Category	*Measure*	*Test*
Spatial	Visualization	Revised Minnesota Form Board Test
	3-dimensional mental rotation	Cube comparison task
	2-dimensional mental rotation	Aerial Orientation Test
	Spatial orientation	Maze task
Attention	Time sharing	Attention sharing
Memory	Short-term duration	Immediate/Delayed Memory[a]
	Long-term stores	Biographical survey
Perception	Perceptual speed	Perceptual speed[a]
	Encoding speed	Encoding Speed[a]
	Vigilance	Perceptual Vigilance[b]
	Pattern recognition	Raven's Advanced Progressive Matrices
Cognitive	Analytic	Analytic Test[c]
	Risk taking	Pilot Decisional Attribute Questionnaire[b]
	Internal timing	Internal timing[a]
	Cognitive complexity	This I Believe test
	Cognitive complexity	O'Conner Abstractness Orientation Scale
	Locus of control	Aviator Locus of Control[b]
	Field independence	Group Embedded Figures Test
	Compulsiveness	Dot Estimation[a]

[a]Present or proposed task adapted from the Basic Attributes Test (Carretta, 1987a). [b]Task adapted from the JAMJET (Jensen, Adrion, Maresh, & Weinert, 1987). [c]Subtest of the Graduate Record Examination (Educational Testing Service, 1988).

Revised Minnesota Form Board Test (RMFBT). The RMFBT (Form BB) was selected to measure visualization (Lickert and Quasha, 1969). This test was designed to measure a person's ability to visualize and manipulate two-dimensional geometric shapes into a whole design. The test consists of 64 two-dimensional diagrams, each comprised of a collection of pieces analogous to an unassembled jigsaw puzzle. For each diagram, the subject was instructed to select from five possible answers the one that correctly represents the pieces put together as a whole figure. The number of correct responses in 20 min was recorded.

Cube comparison task. A cube comparison task, based on a test developed by Ekstrom, French, and Harmon (1976), was used to measure subjects' abilities at mental rotation in three-dimensions. Subjects were presented with 21 drawings of pairs of cubes (similar to children's blocks) that had a unique letter or number on each face of the cube. The task was to determine if the two drawings could represent the same block by mentally rotating the blocks so that they would have the same orientation. The number of correct responses was recorded.

Aerial Orientation Test. This task is based on the Aerial Orientation Test developed in 1947 for the Air Force (Guilford & Lacey, 1947) and was designed to measure a subject's ability to mentally rotate a two-dimensional aircraft outline. The subject's task was to select from five aircraft outlines presented at various rotations the one that showed the same side as a presented aircraft. The test consists of 30 items. Time to complete the test and accuracy were recorded.

Maze task. A maze task was included to measure subjects' abilities at spatial orientation on a fixed map. The task consists of four three-dimensional mazes. A practice maze was given to each subject to familiarize him with the task. Total time to reach the endpoint of each maze successfully was recorded.

Attention Tests

One test was selected to measure an individual's ability to time-share across tasks.

Attention sharing. This test measured a subject's ability to perform dual tasking. The test consists of a two-dimensional tracking task coupled with a digit cancellation task, which was based upon earlier experiments

(Carretta, 1987b; Damos, 1978; North and Gopher, 1976). The test was completely computerized and consisted of 10 practice trials, followed by three subtests, each consisting of three 1-min trials. A tracking task was presented during all three subtests based on a random-order, sinusoidal, rose-petal forcing function. Task difficulty ranged in value from 1 (easiest) to 10 (hardest). To keep the tracking task at the same level of perceived difficulty for all subjects across the testing period, the program automatically increased or decreased tracking difficulty to keep tracking errors at a constant prescribed level. Thus, the tracking task was kept at optimum difficulty for each subject throughout the experiment.

In the digit cancellation task, a digit appeared at a random interval of between 5 and 15 sec. If subjects did not respond to the digit within 4 sec after its presentation, the tracking circle disappeared, forcing the subjects to cancel the digit in order to resume tracking. The tracking task was presented during all three subtests. In the first subtest, subjects canceled one of two digits (1 or 2) on the screen by pressing the corresponding key on a keyboard. In the second subtest, subjects canceled one of eight digits (1 to 8). In the third subtest, subjects performed only the tracking task. Response time (RT) to cancel the digit, distance error for the tracking task, and average level of tracking difficulty were recorded.

Memory Tests

One test was selected to measure an individual's short-term memory capacity. In addition, demographic data was collected via a biographical survey to obtain some measure of the subject's level of experience that could contribute to long-term memory stores.

Immediate/Delayed Memory. This test is based upon the Immediate/Delayed Memory test developed for use in the BAT (Carretta, 1987a). The test is computerized, and consists of two subtests, during which a series of one-digit numbers are flashed on a computer screen for 0.5 sec. The interstimulus interval was 2 sec for half of each subtest (immediate) and 5 sec for the remainder (delayed). In the first subtest, subjects were asked to respond via the keyboard with the number that appeared immediately prior to the number on the screen. In the second subtest, subjects were asked to respond with the number that appeared two numbers prior to the number displayed on the screen. For both subtests, 10 practice trials and 50 test trials were conducted. Accuracy and RT for each subtest were recorded.

Biographical survey. A biographical survey was developed to address the possibility that individual factors such as pilot experience or age may be related to SA. Items included in the survey were age, years of flight experience, number of flight hours, and combat experience. Although these are probably fairly weak measures, they may provide some indication of the degree of relevant long-term memory stores acquired.

Perception Tests

The tests compiled to measure perceptual abilities included tests of perceptual speed and vigilance, encoding speed, and pattern recognition.

Perceptual speed. The perceptual speed task was adapted from an early experiment used to measure perceptual speed (Sperling, 1960) that is being considered for inclusion in the BAT (Carretta, personal communication, January 25, 1989). This test is computerized and consists of a practice session of 10 trials and five subtests of 16 trials each, with stimulus presentation times of 500, 400, 300, 200, and 100 ms, respectively. A three- to seven-digit number was presented on the computer screen for the prescribed stimulus time. After a 500-ms delay, a second number was presented. Subjects had to decide as quickly as possible whether the two numbers were the same or different (due to transposed digits) by pressing designated keys on the keyboard in response. Accuracy and RT were recorded.

Encoding Speed. The encoding speed task was adapted from Posner and Mitchell (1967) and the BAT (Carretta, 1987a). The test is computerized, consisting of three subtests of 32 trials each and 10 practice sessions for each subtest. Subjects were presented with two pairs of letters and had to decide whether the pairs of letters were the same or different. The pairs remained on the computer screen until the subjects responded. Each subtest used a different rule for similarity: physical identity, letters in both pairs must be identical in letter and in case (AA and AA), name identity, both pairs of letters must be composed of the same letter regardless of case (AA and Aa) or categorical identity, and letter pairs need to be either all vowels or all consonants (Ai and Ea). Accuracy and RT were recorded.

Perceptual Vigilance. The Perceptual Vigilance task, designed to measure monitoring and instrument-scanning abilities, was adapted from the JAMJET (Jensen et al., 1987). Subjects were shown a computer screen with 25 rows of 80 red dots on a black background. Subjects were instructed to scan the screen thoroughly for a change in one of the dots from red to magenta that was just above visual threshold. The subject signaled when they noticed the change by pressing any key on the keyboard. Ten trials with stimulus-onset intervals ranging from 1 to 15 sec were administered. Elapsed time from the color change to the subject's response was recorded.

Pattern recognition. The pattern recognition test selected was the Raven's Advanced Progressive Matrices Test (Raven; J. C. Raven, Court, & J. Raven, 1985), which measures a person's ability to perform pattern recognition using nonverbal reasoning skills. The advanced test was chosen over the standard version because it allows greater differentiation of scores at the upper range of performance. Subjects were shown a pattern with a piece missing and were instructed to select the missing piece from eight choices. The test consists of 12 familiarization problems, followed by 36 problems arranged in increasing order of difficulty. The number of correct responses was recorded.

Cognitive Tests

Eight tests were selected to measure higher order cognitive abilities: analytical reasoning, risk-taking tendencies, internal timing, cognitive complexity (2 tests), locus of control, field independence, and compulsiveness/decisiveness.

Analytic subtest. The test used to measure analytical reasoning is a subtest of the Graduate Record Examination (GRE; Educational Testing Service, 1988). The subtest consists of 25 questions with a time limit of 30 min. The test was designed to measure a subject's ability to understand a given structure of arbitrary relations among the presented items and to deduce new information from the relations given. For each question, there are five choices from which subjects are to select the correct response. The number of correct responses was recorded for each subject.

Risk taking. A risk-taking test called the Pilot Decisional Attribute Questionnaire (PDAQ) was adapted from the JAMJET (Jensen et al., 1987). The PDAQ is administered via paper and pencil and was designed to measure a subject's tendencies toward five hazardous attitudes: anti-authority, impulsivity, invulnerability, machismo, and resignation. The test consists of 10 situations and five possible actions for each situation. Subjects were instructed to rank the actions on a 5-point scale ranging from *least likely* (5) to *most likely* (1). Based on these answers, a total score was computed for each of the five attitudes. The attitude receiving the highest score indicates the most likely tendency for each subject.

Internal timing. This test is a modification of a task being considered for inclusion in the BAT (T. Carretta, personal communication, January 25, 1989). Subjects viewed three labeled points (A, B, and C) placed in a straight line on a computer screen. When subjects depress the space bar on the keyboard, a target begins moving from A at a constant velocity and is blanked from the screen as it passes B. The subject's task was to determine when in time the target would reach C and to press the space bar at that time in response. The task used five different velocities and five different point locations. There were 10 practice trials and 50 test trials. The distance between the target's current location and C was recorded for each trial, and feedback was provided.

Cognitive complexity. Two tests measured cognitive complexity: the "This I Believe" Test (TIB; Harvey, 1966) and the Abstract Orientation Scale (AOS; O'Conner, 1971). The TIB consists of five statements beginning with "This I believe about . . . " (Harvey, 1966). The subjects were given 2 min to write two or more sentences concerning their personal beliefs about the topics. These beliefs were analyzed to determine the subject's cognitive complexity level. (The scoring for this test was done by two trained individuals.)

The AOS (O'Conner, 1971) was also included to measure cognitive complexity. The test consists of 30 questions that are answered on a 6-point rating scale ranging from *strongly disagree* (1) to *strongly agree* (6). Only 18 questions were used in the analysis. (The remainder of the questions are distracter questions and were not analyzed.) There was no time limit on the test.

Field independence. Field independence was measured by the Group Embedded Figures Test (GEFT; Oltman, Raskin, & Witkin, 1971). Each problem in the test consists of a complex geometric pattern that contains one of eight simple geometric figures that are presented to the subject. Subjects were required to trace with a pencil the simple geometric figure that is embedded in each complex pattern. Three minutes are provided for a practice section with seven problems, and 5 min for each of two test sections with eight problems. Accuracy and time to complete each section were recorded.

Locus of control. Locus of control was measured by the Aviator Locus of Control subtest, which is a JAMJET modification of the Rotter Internal–External Control Scale for the pilot population (Jensen et al., 1987; Rotter, 1966). The test measures the extent to which a person is internally or externally controlled. The test consists of 29 questions (of which 4 are distracter questions and were not scored), for which the subjects selected which of two statements (an internally controlled response or an externally controlled response) best represented their perception of the world.

Compulsiveness/decisiveness. A dot-estimation task adapted from the BAT was designed to measure compulsiveness/decisiveness (T. Carretta, personal communication, January 25, 1989). The test is computerized and consists of 50 test trials with no practice trials. During the test, subjects are shown two equal-sized square boxes on a computer screen. The boxes contained a number of white dots, with one of the boxes containing one more dot than the other. The subject's task was to indicate as quickly as possible which box contains more dots, using designated keys on the keyboard. Accuracy and RT were recorded for each trial.

SITUATION AWARENESS MEASUREMENT

To determine whether these attributes do indeed relate in any consistent manner to abilities in SA, as opposed to other flight skills, it was also necessary to obtain a measure of subject SA. The Situation Awareness Global Assessment Technique (SAGAT; Endsley, 1987, 1988c) was selected as a measure of SA for this purpose. SAGAT provides an objective measure of SA in manned simulations of the task environment. Using SAGAT, the simulation is stopped at random times and subjects are queried using a computerized tool to determine their SA at that particular point in time. Subjects' answers are compared with the correct answers, which have been simultaneously collected through the simulation computer. The comparison of the real and perceived situation provides an objective measure of SA. The random-sampling method assures that an unbiased measure is obtained across trials and conditions. The SAGAT battery includes 40 queries covering all aspects of SA, including ownship status, the status of other aircraft and ground threats, comprehension, and projection elements.

In addition to possessing a high degree of face validity, the SAGAT technique has been tested in several studies which demonstrated the technique's (a) empirical validity (Endsley, 1989a, 1990b) in that freezing the simulation did not impact subject performance and subjects were able to reliably report SA knowledge for up to 6 min after a freeze without memory decay problems, (b) predictive validity (Endsley, 1990a) by linking SAGAT scores to subject performance, and (c) content validity (Endsley, 1990b) by showing appropriateness of the queries used (for an air-to-air fighter cockpit).

METHOD

The study focused on determining which of the attribute measures discussed might have some relation to SA for military pilots in the single-seat, air-to-air fighter environment. The study consisted of two parts: the collection of

SAGAT data to measure pilot abilities at SA, and the administration of the SA attribute battery.

Subjects

Twenty-five male subjects participated in the SA measurement portion of the study. Of the 25 subjects, 21 were available to participate in the attribute-measurement portion of the study. All subjects were experienced, former military fighter pilots employed by Northrop. The mean subject age was 45.16 years (range = 32 to 68 years). They had an average of 3,582 flight hr (range = 975 to 7,045 flight hr) and an average of 16.9 years (range = 6 to 27 years) of military flight experience. Fourteen of the subjects had combat experience. Sixteen of the subjects were former Air Force pilots, 5 were former Navy pilots, and 4 were former Marine Corps pilots. All subjects reported experience in a multitude of aircraft ($M = 4$ different aircraft, range = 1 to 11). The most frequently mentioned aircraft were the F-4 (72%), F-5 (48%), T-38 (40%), F-100 (24%), T-37 (24%), A-4 (20%), and F-15 (16%).

SA Measurement

Procedure. A set of air-to-air engagements was conducted in a real-time, manned, multiengagement simulator facility. A fighter-sweep mission with a two-fighter (Blue team) versus four-fighter (Red team) force ratio was used for the trials. The objective of the Blue team was to penetrate Red territory, maximizing the kills of red fighters while maintaining a high degree of survivability. The Red team was directed to fly around their assigned combat air patrol (CAP) points until a Blue target was detected in Red airspace. They were then allowed to leave their CAP point to defend against the Blue team. In all cases, specific tactics were at the discretion of the individual pilot teams.

Five teams of six subjects participated in the test. (Some subjects participated on more than one team.) At random points in each trial, the simulator was "frozen" and SAGAT data immediately collected from all six participants. Stops occurred either one, two, or three times during the trial. In some trials there were no stops. At each freeze, the trial was resumed after the specified period for collecting SAGAT data had elapsed (which varied between 30, 60 or 120 sec) and was continued until specified criteria for completion of the mission were met. Subjects completed as many queries as they could during each stop. Queries were presented in a random order. SAGAT data was collected from each team member 36 times over a period of 24 trials for each team.

Prior to conducting the study, all subjects were trained on the use of the simulator, the displays, aircraft handling qualities, and SAGAT. In addition to three instructional training sessions on using SAGAT, each subject participated in at least 18 practice trials during which SAGAT was administered. (Most subjects also had received a substantial amount of training in the simulator in the past.) Thus, the subjects were well trained prior to testing.

Apparatus. Northrop's Integrated Simulation and Systems Laboratory (ISSL) was used for the test. ISSL is a high-fidelity, real-time, interactive, man-in-the-loop simulation facility. ISSL incorporates a Gould mainframe computer which controls simulations and drives Silicon Graphics–generated high-resolution color graphics displays. This test used six manned stations, each configured to represent hypothetical future-generation aircraft. (Hypothetical performance characteristics, weapons, and avionics capabilities were used to keep the simulation at a unclassified level.) The manned station included a simulated head-up display, a tactical-situation display, radar and system controls operated by a touch screen or stick, and throttle control switches. A realistic stick and throttle provided primary flight control.

Attribute Battery Administration

Administration of the attribute battery typically took each subject three test sessions of approximately 2 to 3 hr each to complete. Tests were given in the same order for each subject. The paper-and-pencil tests were given first, followed by the computerized tests.

Apparatus. The computerized tests were administered via an AST 386 personal computer with an 18-in. NEC Multisync color monitor and a Kraft KC3 joystick. The data was electronically stored and transferred to a Macintosh computer for analysis.

RESULTS

SA Stability

SAGAT data collected from the subjects were compared to actual values for each measurement at each stop. Only subjects' knowledge of the location of enemy aircraft was analyzed for this study. From this, an average SA score for each subject across the 36 measurement points was calculated. SA scores varied from .038 to .330 across individuals (0 = no SA, 1 = perfect SA).

An analysis of variance (ANOVA) was performed to determine whether the SA scores were independent of the team (1 to 5) and side (Red vs. Blue) to which each subject was assigned. The results of the ANOVA revealed that neither team, $F(4, 24) = .842$, nor side, $F(1, 24) = .071$, were significantly related to SA score at the .05 level, indicating that these factors did not significantly impact subject SA scores.

The second question to be addressed in this study was whether SA abilities are stable within a given individual. To investigate this, SA scores for the three individuals who participated in the study twice (once on each of two teams) and the one individual who participated in the study three times (once on each of three teams) were further evaluated. Test–retest reliability scores calculated for each individual subject were .99, .92, .98 and .98, respectively, indicating a fairly high level of stability for SA within subjects. The results of these two analyses support the hypothesis that there are fairly constant individual differences in SA.

SA Correlates

Of the 25 participants in the study to collect SAGAT data, 21 were available to complete the attribute test battery. The attribute tests were scored for number correct, RT, and/or number of errors, as appropriate. Because some attribute tests consisted of more than one subtest, a total of 31 variables were examined. SA scores for each subject were correlated with scores on each variable using a Pearson pairwise correlation matrix (because all subjects were not available for every test). The results of this analysis are presented in Table 2, along with means, standard deviations, and Pearson's *r*s for each variable.

Spatial. Three of the four spatial tests showed a moderate correlation with SA: the RMFBT ($r = .317$), the cube comparison test ($r = .353$), and the maze task ($r = -.354$). Although the AOT was similar to the cube comparison task in requiring mental rotation of objects, it had a fairly low correlation with the SA measure ($r = .150$). This may be due to the fact that the task was too simple for pilots, or that some of the subjects discovered that it was possible to perform the task in a way that did not require spatial rotation. In general, it would seem that these results provide evidence for the relation between spatial skills and SA.

Attention. The time-sharing task provided some rather confusing results. Although RTs on neither the two-digit nor the eight-digit cancellation tasks showed very high correlations with SA (rs = –.138 and –.250, respectively), the level of difficulty reached in the tracking task was highly correlated with SA (r = .717). Although it is not immediately obvious why a psychomotor task should be related to SA, an explanation can be hypothesized. Most likely, those pilots who possess very good tracking skills are able to devote much more of their attention towards the assessment of the situation instead of towards manually flying the aircraft, thus they would have higher SA. If this is true, however, it would be expected that RT on the digit cancellation tasks would have reflected this spare capacity as shown by a higher correlation with SA. It is possible that either (a) the subjects were using another strategy besides attention sharing (such as rapid switching between tasks), or (b) the digit cancellation tasks were too simple to provide the level of sensitivity needed. A revision to this task will be necessary to develop any firm conclusions regarding the role of attention sharing in SA abilities.

Memory. The Immediate/Delayed Memory test did not support the hypothesized relation between short-term memory and SA. Although overall RT was moderately correlated with SA (r = .389), it was not in the expected direction. That is, those subjects who on average took longer to respond with the correct answer had higher SA scores. Number of errors on this task was not highly correlated (r = –.071) with SA. To investigate this finding, the test procedure was examined. In the test, the subject is asked to respond to the presentation of a stimulus with a prior stimulus. Thus, it taps a subject's ability to hold competing stimuli in memory, despite the presence of new stimuli. It is possible that this skill may be an actual disadvantage in the rapidly changing environment of the combat pilot. Or, possibly, the result is simply a fluke. More data are probably needed on the subject. In addition, it may be that an alternate measure of short-term memory, such as memory span, may be more appropriate.

Four biographical measures—subject age, years of flight experience, number of flight hours, and combat experience—were also evaluated as possible indices of long-term memory stores. Although the hypothesized relation between these measures and long-term memory stores is weak (as number and quality of stores are not necessarily related to time per se), no better independent measures are known. It is not greatly surprising, therefore, that these measures were not highly correlated with SA. As a matter of fact, all four showed a small negative correlation (rs = –.225, –.233, –.304, and –.164, respectively), indicating that the younger pilots who had less flight experience and no combat experience (which were highly intercorrelated) actually scored slightly better on the SA measure. As all of the subjects were well experienced (≥7 years), it is highly likely that the number of years or hours of experience would cease to be as important to SA as the quality of time spent and other individual attributes.

TABLE 2
Correlations Between Attribute Measures and Situation Awareness

Category	*Measure and Variable*	*n*	*M*	*SD*	*Pearson's r*
Spatial	Revised Minnesota Form Board Test				
	Number correct	21	42.76	9.42	.317
	Cube comparison task				
	Number correct	19	12.90	4.51	.353
	Aerial Orientation Test				
	Number correct	20	65.15	3.44	.150
	Maze task				
	Average test time (in seconds)	20	105.12	44.66	−.354
Attention	Attention sharing				
	Two-digit cancelation RT	15	1.36	0.26	−.138
	Eight-digit cancelation RT	15	1.43	0.16	−.250
	Tracking-only task difficulty level	15	4.29	0.75	.717
Memory	Immediate/Delayed Memory				
	Total test RT	13	1.05	0.45	.389
	Total errors	13	2.73	2.63	−.071
	Biographical survey				
	Age	21	44.28	9.18	−.225
	Experience				
	Years	21	16.90	6.26	−.233
	Flight hours	21	3619.28	1551.28	−.304
	Combat	21	0.52	0.51	−.164
Perception	Perceptual speed RT				
	Subtest				
	1	14	1.18	0.18	−.041
	2	14	1.07	0.17	−.167
	3	14	1.07	0.15	−.007
	4	14	0.98	0.12	.066
	5	14	0.94	0.13	−.448
	Total errors	14	9.43	1.56	.366
	Total test RT	14	1.05	0.14	−.128
	Encoding Speed RT				
	Physical subtest RT	14	0.93	0.17	−.074
	Name subtest RT	14	0.99	0.18	−.295
	Categorical subtest RT	14	1.53	0.32	−.547
	Total errors	14	2.57	1.60	−.264
	Perceptual Vigilance				
	RT	18	2.92	3.47	.041
	Raven's Advanced Progressive Matrices				
	Number correct	20	22.15	4.82	.243
Cognitive	Analytic subtest				
	Number correct	21	14.05	4.11	.073
	Risk taking				
	Predominant attitude	20	–	–	–
	Internal timing				
	Average absolute error	14	62.99	22.68	−.074

(Continued)

TABLE 2
(Continued)

Category	*Measure and Variable*	*n*	*M*	*SD*	*Pearson's r*
	This I Believe test	21	4.00	0.00	–
	O'Conner Abstractness Orientation Scale	21	4.00	0.00	–
	Aviator Locus of Control	21	–	–	–
	Group Embedded Figures Test				
	Number correct	21	15.76	3.46	.385
	Dot Estimation				
	Total test time	17	591.06	322.03	–.418
	RT	17	11.06	5.82	–.382
	Number correct	17	45.94	8.99	–.415

Perception. Of the five perceptual speed subtests, only RT on Subtest 5, the shortest presentation time, showed a moderate correlation with SA ($r = -.448$). It is likely that the longer presentation times of the other subtests were not sufficiently demanding to pick up on any true perceptual differences between subjects. In addition, number of errors on the perceptual speed task was moderately correlated with SA ($r = .366$). Although it is understandable that subjects with fast perceptual speed would have better SA, it is interesting that even those with more errors had higher SA scores. Of the three Encoding Speed subtests, only RT on the Categorical subtest was highly correlated with SA ($r = -.547$). Similarly, this subtest taps the most difficult level of coding, and is probably the most sensitive to individual differences in encoding.

Neither of the other two perceptual tasks, the Perceptual Vigilance subtest or the Raven were very highly correlated with SA (*r*s = .041 and .243, respectively). Although vigilance may be an important skill in many flying tasks, it is highly likely that it was of minimal importance in the air-to-air combat simulation, which formed the basis for the SA measure in this study. It is therefore not surprising that no correlation was found. It is interesting, however, that the Raven, as a measure of pattern recognition, was not very highly correlated either. It should be noted that the Minnesota Form Board Test and the GEFT, although they purport to measure other skills, are both highly correlated with the Raven (*r*s = .576 and .533, respectively) and with each other ($r = .576$). Although slightly different, all three tests seem to tap into a subject's ability to discern patterns in complex figures. The correlation of the Raven with the SA measure was somewhat low, but both of the other tests had higher correlations (*r*s = .317 and .385, respectively). It would seem, therefore, that the issue of pattern matching as it relates to SA needs to be studied further. Perhaps in the future, use of the standard version of the Raven would be preferable to the advanced version, which was developed for gifted children.

Cognitive. Neither the Analytic subtest of the GRE nor the internal timing task (*r*s =.073 and –.074, respectively) were highly correlated with the SA measure used in this study. It is likely that such abilities are far more important to the higher levels of SA (comprehension and projection of future scenarios) than they are to knowledge of other aircrafts' location. The PDAQ produced a characteristic risk-taking attitude for each of the subjects. The majority of the subjects fell into the "invulnerable" category. Due to insufficient variance on this variable, no assessment of its relation to SA could be made.

The TIB and the OAS both categorized all 21 of the subjects as "concrete thinkers." Because there was no variance on this variable, no assessment of its relation to SA could be made. The Aviator Locus of Control test found all but one of the subjects to be internally controlled. Again, due to lack of variance, no real assessment could be made of its relation to SA.

The GEFT, a measure of field independence, did show a moderate correlation with SA ($r = .385$), indicating that those subjects who were more field independent were better at SA. As already mentioned, however, due to the high correlation of the GEFT with the RMFBT and the Raven, it is difficult to say just which abilities are really being tapped.

Last, the Dot Estimation task also showed a moderate correlation with SA for RT and number correct (*r*s = –.382 and –.415). As with the perceptual speed test, subjects who responded faster had better SA, even though they may have made more errors. Considering that RT on this task was correlated at .459 with RT on the perceptual speed test and that number correct was correlated at –.492 with number of errors on the perceptual speed test, this is not surprising. Even though the Dot Estimation task is supposedly a measure of compulsiveness/decisiveness, as taken from the BAT, it may be that the test also taps a subject's perceptual skills to some degree. It may also tap into spatial skills, as evidenced by a fairly high correlation between RT on this task and scoring on the RMFBT ($r = -.467$).

CONCLUSIONS AND RECOMMENDATIONS

In summary, this study supports the hypothesis that some individuals are better at SA than others. This ability seems to be fairly consistent within individuals. It also supports the importance of spatial and perceptual skills for SA. Partial support was provided for the importance of attention-sharing and pattern-matching skills, although more data are probably needed to reach any firm conclusions. No support was provided for the hypothesized relation between SA and memory or analytical skills. Further research on these issues using other measures of the attributes and SA is recommended. Although both field independence and impulsiveness/decisiveness were related to SA, it is difficult to draw any firm conclusions on these measures due to their high intercorrelation with other measures.

Some recommendations can be made from these results on methods to enhance aircrew SA. There is some evidence that some of the attributes examined in this study can be improved through training. Time sharing (Damos, 1978; Damos & Wickens,1980; Gabriel & Burrows, 1968) has been found to be improved through training. Spatial skills, unfortunately, have generally not been found to be trainable (Levine, Schulman, Brahlek, & Fleishman, 1980), and therefore will need to be identified in the selection process. Information is needed on the potential trainability of perceptual skills, pattern matching, field independence, and compulsiveness/decisiveness.

As the first to attempt to examine these issues in depth, this study was fairly limited, and its generalizability should be view with three caveats in mind. First, it is based on a fairly small group of experienced pilots. The subjects included in this study can be assumed to have skills that are above certain thresholds for satisfactory performance in the flight environment. Very different results might be obtained if all those entering flight school or the general population were to be examined. Second, this study is based on pilot SA in an air-to-air fighter-sweep mission. Different skills might be called on by pilots in other aircraft in performing other kinds of tasks. For instance, vigilance may be much more important for a bomber or commercial pilot. Third, the SA measure used here gauged only the subject's knowledge of other aircraft's locations. Although this is a crucial component of SA for the fighter pilot, upon which most other components hinge, other components of SA (e.g., projection of future scenarios) may draw on other skills.

In general, it is recommended that this effort be expanded to a larger group of subjects, preferably including those who have not yet completed flight training and those who are currently pilots in a wide variety of aircraft and missions. Several components of SA should be examined to provide a more well-rounded dependent measure. In addition, the previously noted modifications should be made to the test battery so that a better assessment of subjects' skills is acquired.

Overall, this study examined the heretofore-theorized belief that certain individuals are superior in their ability to acquire and maintain SA. Support was provided for this concept, and an initial examination was made of the attributes that may lead to these individual differences. The use of selection and focused training programs to enhance these skills may lead to a population of pilots who have superior capabilities in achieving this very crucial construct.

ACKNOWLEDGMENTS

This research was sponsored by the Northrop Corporation.

We express our appreciation to Dr. Thomas Carretta, Dr. Richard Jensen, and CMDR Michael Pianka for their invaluable assistance in creating the test battery. Thanks are also due to Sam Rodriquez for his marvelous programming, to Chuck Semple for his support and assistance in completing this project, and to Drs. Michelle Robertson and Hal Hendrick for their assistance in scoring the TIB test.

REFERENCES

Bolstad, C. A. (1991). Individual pilot differences related to situation awareness. In *Proceedings of the 35th annual meeting of the Human Factors Society* (pp. 52–56). Santa Monica, CA: Human Factors Society.

Bolstad, C. A., & Rodriquez, S. L. (1991). *Evaluation of pilot attribute measures* (NOR DOC No. 90-01). Hawthorne, CA: Northrop Corporation.

Carretta, T. R. (1987a). *Basic Attributes Test (BAT) system: Development of an automated test battery for pilot selection* (Rep. No. AFHRL-TR-87-9). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Carretta, T. R. (1987b). *Time sharing ability as a predictor of flight training performance* (Rep. No. AFHRL-TR-86-69). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Damos, D. L. (1978). Residual attention as a predictor of pilot performance. *Human Factors, 20,* 435–440.

Damos, D. L., & Wickens, C. D. (1980). The acquisition and transfer of time-sharing skills. *Acta Psychologica, 6,* 569–577.

Educational Testing Service. (1988). *Graduate Record Examination practice manual.* Princeton, NJ: Author.

Ekstrom, R. B., French, J. W., & Harmon, H. H. (1976). Cube comparison test. In *Manual for kit of factor referenced cognitive tests* (pp. 150–154). Princeton, NJ: Educational Testing Service.

Endsley, M. R. (1987). *SAGAT: A methodology for the measurement of situation awareness* (NOR DOC 87-83). Hawthorne, CA: Northrop Corporation.

Endsley, M. R. (1988a). *A construct and its measurement: The functioning and evaluation of pilot situation awareness* (NOR DOC 88-30). Hawthorne, CA: Northrop Corporation.

Endsley, M. R. (1988b). Design and evaluation for situation awareness enhancement. In *Proceedings of the 32th annual meeting of the Human Factors Society* (pp. 97–101). Santa Monica, CA: Human Factors Society.

Endsley, M. R. (1988c). Situation Awareness Global Assessment Technique (SAGAT). In *Proceedings of the National Aerospace and Electronics Conference (NAECON)* (pp. 789–795). New York: IEEE.

Endsley, M. R. (1989a). A methodology for the objective measurement of situation awareness. In *Situational awareness in aerospace operations* (pp. 1/1–1/9). Copenhagen, Denmark: North Atlantic Treaty Organization, AGARD.

Endsley, M. R. (1989b). Pilot situation awareness: The challenge for the training community. *Proceedings of the Interservice/Industry Training Systems Conference (I/ITSC)* (pp. 111–117). Fort Worth, TX: American Defense Preparedness Association.

Endsley, M. R. (1990a). Predictive utility of an objective measure of situation awareness. In *Proceedings of the 34th annual meeting of the Human Factors Society* (pp. 41–45). Santa Monica, CA: Human Factors Society.

Endsley, M. R. (1990b). *Situation awareness in dynamic human decision making: Theory and measurement* (NOR DOC 90-49). Hawthorne, CA: Northrop Corporation.

Gabriel, R. F., & Burrows, A. A. (1968). Improving time sharing performance of pilots through training. *Human Factors, 10,* 33-40.

Gopher, D., & Kahneman, D. (1971). Individual differences in attention and the prediction of flight criteria. *Perceptual and Motor Skills, 33,* 1335–1342.

Griffin, R. (1988). *An improved automated selection system for navy pilots* (Rep. No. DTIC-AD-A203438). Pensacola, FL: Naval Air Station, Naval Aerospace Medical Laboratory.

Guilford, J. P., & Lacey, J. I. (1947). *Printed classification tests* (Army Air Force Aviation Psychology Program Research Report No. 5). Washington, DC: U.S. Government Printing Office.

Harvey, O. J. (1966). System structure, flexibility and creativity. In O. J. Harvey (Ed.), *Experience, structure and adaptability* (pp. 39–65). New York: Springer.

Jensen, R. S., Adrion, J., Maresh, J., & Weinert, C. (1987). *Pilot judgment skills test battery (Final report).* Columbus, OH: Aviation Research Associates.

Kantor, J. E., & Bordelow, V. P. (1985, March). The USAF pilot selection and classification program. *Aviation, Space and Environmental Medicine, 56,* 258–261.

Kass, S. J., Herschler, D. A., & Companion, M. A. (1990). Are they shooting at me?: An approach to training situation awareness. In *Proceedings of the 34th annual meeting of the Human Factors Society* (pp. 1352–1356). Santa Monica, CA: Human Factors Society.

Kirsch, H. (1976, November). *Strategy of selection and psychological tests as part of the screening for flight trainee applicants at DLH.* Paper presented at the fourth general flight crew training meeting of the International Air Transport Association, Lucerne, Switzerland.

Levine, J. M., Schulman, D., Brahlek, R. E., & Fleishman, E. A. (1980). *Trainability of abilities: Training and transfer of spatial visualization* (Defense Technical Information Center Rep. No. AD-A283497). Washington, DC: Advanced Research Resources Organization.

Lickert, R., & Quasha, W. H. (1969). *Revised Minnesota Paper Form Board Test.* New York: Psychological Corporation.

North, R. A., & Gopher, D. (1976). Measures of attention as predictors of flight performance. *Human Factors,* 18, 1–14.

O'Conner, J. (1972). Developmental changes in abstractness and moral reasoning. (Doctoral dissertation, George Peabody College for Teachers). *Dissertation Abstracts International, 32,* 4109A.

Oltman, P. K., Raskin, E., & Witkin, H. A. (1971). *Group Embedded Figures Test (GEFT).* Palo Alto, CA: Consulting Psychologists Press.

Posner, M. I., & Mitchell, R. F. (1967). Chronometric analyses of classification. *Psychological Review.* 74, 392–409.

Press, M. (1986). *Situation awareness: Let's get serious about the clue-bird.* Unpublished manuscript.

Raven, J. C., Court, J. H., & Raven, J. (1985). *Raven's Advanced Progressive Matrices.* London: H. K. Lewis.

Robertson, M. M. (1984). Personality differences as a moderator of mental workload behavior: Workload, performance and strain reactions as a function of cognitive complexity. In EDITOR (Ed.), *Proceedings of the 28th annual meeting of the Human Factors Society* (pp. 690–694). Santa Monica, CA: Human Factors Society.

Rotter, J. B. (1966). General expectations for internal versus external control of reinforcement. *Psychological Monographs: General and Applied, 80,* 1–28.

Sperling, G. (1960). The information available in brief presentations. *Psychological Monographs: General and Applied, 74,* 1–29.

Stasz, C., & Thorndyke, P. W. (1980). *The influence of visual–spatial ability and study procedures on map learning skills* (Rep. No. N-1501-ONR). Washington, DC: Office of Naval Research.

Thorndyke, P. W., & Stasz, C. (1980). Individual differences in procedures for knowledge acquisition from maps. *Cognitive Psychology, 12,* 137–175.

Witkin, H. A., & Goodenough, D. R. (1977). *Field-dependence revisited* (Rep. No. RB-77-16). Princeton, NJ: Educational Testing Service.

Witkin, H. A., Goodenough, D. R., & Oltman, P. K. (1977). *Psychological differentiation: Current status* (Rep. No. RB-77-16). Princeton, NJ: Educational Testing Service.

Witkin, H. A., Moore, C. A., Goodenough, D. R., & Cox, P. W. (1977). Field dependent and field independent cognitive styles and their educational implications. *Review of Educational Research, 47,* 1–64.

Youngling, E. W., Levine, S. H., Mocharnuk, J. B., & Weston, L. M. (1976). *Feasibility study to predict combat effectiveness for selected military roles: Fighter pilot effectiveness* (Rep No. MDC E1634). St. Louis, MO: McDonnell-Douglas.