Two measurement methods for situation awareness were tested on 2-person aircrews and compared for the information they yielded on team situation awareness. Forty-one crews of low experience level military aviators flew 2 different scenarios, 1 in a full-mission simulator, a 1 in a low-fidelity trainer. Team situation awareness was measured by instructor and observers in the high-fidelity simulator scenario and by responses to questions on flight knowledge in a scenario in the low-fidelity trainer. Scores on both measures were found to be reliable. Team situation awareness scores based on flight knowledge collected in the low-fidelity scenario were significantly correlated with team performance in the high-fidelity simulation ($r = .41, p < .05$). The 2 team situation awareness scores (1 from the high-fidelity simulation and 1 from the low-fidelity simulation) were also significantly correlated ($r = .43, p < .05$). These findings and related information are discussed in relation to the use of the measurements in situation awareness training for low experience level pilots.
There is widespread agreement that situation awareness is important in aviation. It is crucial in the single-seat fighter cockpit as well as in the multicrew cockpits where a variety of missions are undertaken (Prince & Salas, 2000). Although the majority of military and commercial flights are accomplished with more than one crew member, it is individual situation awareness that has received the most attention. The focus of the research reported here is team situation awareness measurement. The results will be used to lay the groundwork for training low experience level aviators in situation awareness.

Two widely used definitions of situation awareness differ appreciably. One places emphasis on cognitive elements (“Situation awareness is the perception of 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, 1995b, p. 36), whereas the other includes the action taken based on the awareness (situation awareness involves “the pilot’s continuous perception of self and aircraft, in relation to the dynamic environment, threats and mission and the ability to forecast and then execute tasks based on that perception”; Carroll, 1992, p. 5).

Defining team situation awareness is more elusive. One definition of team situation awareness is “the degree to which every team member possesses the situation awareness required for his or her responsibilities” (Endsley, 1989). Because other crew members and their task performance are important to the situation, we suggest that this definition be amended slightly to emphasize team requirements. The definition would explicitly include the term team responsibilities, and would become “awareness that each member has for his or her task responsibilities and his or her team responsibilities.” The team responsibilities include maintaining sufficient knowledge of other team members’ responsibilities so that monitoring and backing up of other team members can occur. There is another definition of team situation awareness that is commonly used in many crew resource management (CRM) training programs and in the military service where this research was conducted. Essentially, team situation awareness is defined by the behaviors that suggest its existence (e.g., identifies problems or potential problems, provides information in advance; Prince & Salas, 1993).

Within the undergraduate rotary-wing community of one military branch, situation awareness was identified as an important crew skill (Stout, Salas, & Fowlkes, 1997), but its training has not been emphasized. Pilots in undergraduate training fly with experienced instructors and concentrate on developing technical skills. At the end of this phase of training, they transition to the position of copilot in a multicrew aircraft. In this position, they can no longer rely on the instructor to maintain awareness for both of them, but must extend their own awareness to the other crew members (their team) both for giving and receiving information.

Effective training for team situation awareness cannot be implemented until a reliable, valid measurement method to diagnose and assess changes in situation awareness is developed. Not only is measurement necessary for evaluation, but it
is required for the feedback that is part of the active training recommended for skill development (Prince & Salas, 1993).

Situation awareness is complex and dynamic and this presents a formidable challenge to developing psychometrically sound measurement techniques. Because situation awareness is “continuously modified and updated over time” (Salas, Prince, Baker, & Shrestha, 1995, p. 124) in response to the demands of a dynamic environment, assessments should incorporate multiple measurements, taken at different periods of the flight.

Two distinctly different methods for measuring individual situation awareness correspond to the definitions given by Endsley (1995b) and Carroll (1992). The Situation Awareness Global Assessment Technique (SAGAT; Endsley, 1988, 1995a) is used as an explicit measure of perception, comprehension, and projection of situation awareness elements (Endsley, 1995a). Explicit measures such as SAGAT use information gathered in flight simulations by stopping those simulations and asking pilots questions about the existing situation (e.g., the status of the plane, the environment, and the mission). Implicit measures (Sarter & Woods, 1995) capture the action taken as a result of situation awareness and thus are useful when measuring awareness defined to include action. Implicit measures use a scenario with built-in probes that provide cues to problems or discrepancies that lead to observable changes in behavior (Sarter & Woods, 1995). Both explicit and implicit measures lend themselves to multiple measurements.

Neither of these two situation awareness measurement methods was designed specifically for assessing team situation awareness. Still, a natural starting point for developing and evaluating a method for team situation awareness measurement is to explore the potential of existing measurement methods. The implicit method appears to be more appropriate for measuring team situation awareness that is defined by crew behaviors because it incorporates actions within the team that contribute to team awareness. The implicit method is used in CRM training with more experienced aircrews. Unfortunately, it has several drawbacks that limit its practical use for the intensive, focused training required for those with far less experience than air-carrier crew members. Whether for training or evaluation, implicit measures are best suited for the high-fidelity simulator where multiple cues with varying degrees of subtlety are available. This is a problem in many training organizations where the high-fidelity simulators are already fully scheduled for technical flight training.

In addition, for accurate evaluation with the implicit measures, instructors or evaluators must be trained to observe, classify, and evaluate the actions of a crew in a busy scenario and they must be calibrated in their ratings. This adds to the time and money required for instructor training. An instructor’s focus is dependent on his or her biases, the crew’s actions in each area (particularly evidence of a problem with procedural knowledge), and the design of the scenario (events included in scenarios can call for different knowledge and skills). Despite widespread use,
therefore, there are apparent problems with implicit measures of team situation awareness.

Low-fidelity trainers have been accepted in many areas of pilot training because they have the advantages of economy and flexibility (Jentsch & Bowers, 1998) and they have been shown to be effective in training specific CRM skill behaviors (Brannick, Prince, & Salas, 2005). However, the possibilities for scenarios where multiple behaviors related to situation awareness can be observed are far more limited than in the high-fidelity simulators. In the most basic trainers, there are no tactile cues and visual cues are minimal; some instruments are missing and some cannot be controlled by the experimenter. This lack of cuing and control of the simulator restricts the opportunity to observe actions based on situation awareness and does not recommend the use of low-fidelity trainers for implicit methods of measurement. Some scenarios have been used in low-fidelity trainers that include problems with tasks beyond flying (e.g., passenger emergency) but such scenarios are not realistic for low experience level pilots.

Although scenario possibilities are far more limited with the low-fidelity simulator, it is able to support explicit measurement methods like SAGAT, where the instructor can easily determine if the crew members have the necessary flight knowledge. It takes no special training in observation and classification on the instructor’s part and the instructor’s evaluation can be limited to the questions developed for the scenario. Such a capability presents serious potential as a measurement method (and eventually a training platform).

Before recommending either measurement method for use in team situation awareness for low experience level aviators, these two measurement methods (implicit and explicit) needed to be compared for the information they provide. Although SAGAT has been used with low-fidelity trainers, it has not been used with realistic flight scenarios for crews who have more than a year’s flight training but are still at a relatively low level of experience. There is an indication that the ability to gain and maintain situation awareness in the cockpit is developed over time (Prince & Salas, 1998) and with experience (Carretta, Perry, & Ree, 1996). This suggests that research results found with nonaviators and with more experienced aviators may not be applicable to aviators with a particular level of experience. To compare the measurement methods for this level of experience, we first obtained permission to conduct research in the training command. We developed two separate scenarios, one for the high-fidelity simulator (for implicit measures) and the other for the low-fidelity trainer (for explicit measurement). Both scenarios required awareness on the part of the crew and each represented part of the training the participants had received. The scenario in the high-fidelity simulator included a more complete sample of that training, was more familiar to the participants, and by definition, was a more faithful representation of the flight situation. In addition to the normal flight sequences from one airfield to another, it contained several emergencies that required recognition of the situation, connecting each to the right
procedures and applying those procedures. Varied, realistic cues were available to the crew members. The scenario supplied multiple opportunities to observe and rate performance as well as situation awareness. The scenario on the low-fidelity trainer was representative of a normal flight from Point A to Point B. Only instruments for basic flight were represented in the trainer. There were no emergencies in the scenario, but there were instructions (e.g., climb and maintain an altitude of 4,000 ft) and there was information about traffic and weather (e.g., multiple thunderstorms moving through the vicinity of the destination airfield) given to the crew by the air traffic controller (ATC).

PURPOSE OF THE STUDY

The primary purpose of the study was to evaluate two methods of measurement for team situation awareness, specifically to compare an explicit method of team situation awareness measurement used with a low-fidelity simulation with an implicit method of team situation awareness measurement made from observations in a high-fidelity simulation and then to compare each with independent ratings of team performance. The secondary purpose was to explore the potential application of our findings to training low experience level aviators in team situation awareness.

Two experienced military instructors were trained to rate the behaviors in the high-fidelity scenario and to give scores for performance and for situation awareness and other team interaction skills (the other interaction skills were used in separate research; Brannick, Prince, & Salas, 2002). Questions were developed for the low-fidelity simulation and answer forms were produced. With these, we set out to answer our research questions:

1. Does each of these team situation awareness measurement methods yield reliable measures?
2. Do the measures of team situation awareness relate to team performance?
3. Can measures of flight knowledge gathered in a low-fidelity scenario distinguish between crews on team situation awareness in a low experience level population?
4. Is there a correlation between the situation awareness as measured by the explicit method and that measured by the implicit method?

The question of reliability of the two measures had to be answered first. Clearly, if the measures were not reliable, then observing significant correlations with any other variable would be precluded. Next, because we have no clear criteria on situation awareness with which to compare our measures, we chose to look at the relation of our measures to a separately derived measure of performance. These performance measures from the high-fidelity simulation were based primarily on objective criteria, such as established emergency procedures. In this case, only the performance in
the high-fidelity scenario could be used because all crews completed the routine flight requirements of the low-fidelity scenario without incident. SAGAT scores have been linked to performance in the scenario in which they were obtained (Endsley, 1990). There is also a suggestion of consistency of SAGAT scores from one scenario to a similar scenario (Endsley & Bolstad, 1994). Our research extends literature to test whether an explicit measurement score is meaningfully related to performance in an entirely different scenario. Although situation awareness changes, even during a flight, if some aviators are able to gain and maintain situation awareness better than others, we would expect overall situation awareness in one flight situation to be at least somewhat related to that in another flight situation.

Our next question was important because the crew members to which the measurement is to be applied are similar in both training and experience. Experience is an important predictor of situation awareness (Carretta et al., 1996). Not only is experience important, but each crew member’s awareness is affected also by training and general cognitive abilities (Carretta et al., 1996). For the pilots in the community that we were studying, experience with the simulator and the aircraft simulated was identical. All had been able to complete the same training program at the same time. They had been selected for aviation training from a large pool of applicants and then selected again after initial fixed-wing flight training for the rotary-wing community. In addition, they were a more homogeneous group than the general population. The sample that volunteered for this research were all in their 20s, all were college graduates, all aspired to be professional pilots, and all but 3 were male.

Our last question was to see whether the two different methods for measuring situation awareness were correlated. Although both methods purport to be measuring situation awareness, they each take a different perspective. The explicit measurement method appears to measure potential team situation awareness; that is, what each individual knows on his or her own, but not what of that is contributed to the team. In fact, it measures flight-relevant knowledge that is believed to be associated with situation awareness. Examples of flight-relevant knowledge include current position, assigned altitude, the relative position of traffic, considerations involved in possible landing sites, and the implications of nearby weather. The implicit method does not provide any direct information about flight knowledge but only information about the actions that may have resulted from that knowledge. If actions are inhibited for any reason, it is only a partial measure of team situation awareness and may well underrepresent the cognitive aspects of situation awareness.

METHOD

Participants

Military pilots who recently had completed flight training, first in fixed-wing aircraft and then in basic helicopter flight and instrument training in the ro-
tary-wing aircraft, volunteered for the research. Participants were randomly assigned to two-person teams, with the restriction that each team contained pilots who had not flown together previously. The leadership role in each crew was determined by a coin toss. Forty-eight crews flew both scenarios. There were equipment malfunctions with the low-fidelity trainer for 5 crews and videotapes of 2 crews were used to train instructors. As a result, the total number of crews whose data was used in the research was 41.

Simulations

**High-fidelity flight simulation.** A flight instructor designed the scenario for the full-mission helicopter flight simulator. The mission was to pick up some airplane parts at an airfield about 150 miles from the crews’ base. There were several problems in the scenario: icing conditions at 6,000 ft, a fuel boost pump failure, and an electrical fire that forced termination of the flight at an alternate airfield.

**Low-fidelity flight simulation.** A pilot proficient in developing training scenarios created the script. In this scenario, the crew members flew to a small town on the east coast of Florida, a distance of about 60 miles from their home base. This was a routine flight with no equipment malfunctions and no need for emergency procedures.

We deviated from traditional SAGAT procedure in three ways that were necessitated by the equipment and the experience level of the aviators. First was the mode of presentation of the question: Instead of using the computer, questions were given on a sheet of paper. Second, the scenario script differed from SAGAT, where multiple trials of similar situations are flown. A single flight from one airfield to another was more realistic for the participants in this research. The scenario developer and two other pilots generated 48 questions that crews should be able to answer at three selected stopping points in the flight. These questions were asked at the same stopping points for all crews (to ensure the situations would be equal for all participants) rather than at the traditional random points. Because 11 questions relevant to situation awareness throughout the flight were repeated at different stopping points, we added another question to our research: Do crew members improve performance on repeated questions, even though they are asked at a different time and are likely to have different answers?

Questions were intended to tap factual knowledge of crews who were properly aware of the situation (based on expert evaluation of the situation). Most questions in each of the three question sets concerned the part of the flight that preceded that question set (from the brief to the first stop for Question Set 1, from the first stop to the second stop for Question Set 2, and from the second stop to the last stop for Question Set 3) although each question set contained some questions that pertained to the other parts of the flight. Next, six senior military pilots who had not participated in the sce-
nario development viewed a videotape of a crew flying the scenario and reviewed the
questions for each stopping point. They agreed that the questions tapped knowledge
that crews with good situation awareness would possess.

Procedure

Each crew flew in the full-motion simulator first. The flight portion was video-
taped for subsequent evaluation, and the preflight briefing sessions were re-
corded on audiotape. After debriefing the flight from the full-motion simulator,
the pilots had a 15-min break followed by the scenario on the low-fidelity
trainer. All crews read a prepared brief. After the flight began, the scenario was
frozen and screens blanked at the three predetermined times. Questions were
presented to each crew member immediately after the screen blackout. Pilots
were given 2 min to read and answer questions individually. When the papers
were collected, the flight was resumed. Following this simulation, pilots were
debriefed, thanked, and allowed to resume their normal duties.

Performance Evaluation Measures

Two forms, one problem oriented and the other process oriented, were used to
evaluate team behavior in the full-motion flight scenario. The problem-oriented
form provided relatively objective evaluations of specific events. Included were
each crew member’s handling of the icing, boost pump, and electrical fire prob-
lems according to established procedures and their scores on each of these prob-
lems as a team. The process-oriented evaluations were on situation awareness
and six CRM dimensions (Prince & Salas, 1993). These six CRM dimensions
were part of a separate research project (Brannick et al., 2002).

Instructor Ratings for the High-Fidelity Simulation

Two experienced instructor pilots served as judges for the performance evalua-
tions in the full-motion simulation. Both judges completed a 3-day training ses-
son, in which they reviewed the importance of team coordination and the defi-
nitions of each skill dimension. The judges reviewed several tape segments so
that they could practice performance evaluations and record specific behaviors.
Judges independently assigned behaviors to skill categories. During this phase
of training, the judges were given feedback about the appropriateness of their
observations and evaluations, and the rating forms were revised. Both judges
then watched two full-length simulation tapes, using revised versions of the
evaluation forms. Because both instructors were experienced and were familiar
with the required technical responses of the crews to each of the emergencies,
they did not need training in problem-oriented performance evaluation.
Only two videotapes were used for training because the tapes used to train the judges could not be used subsequently. The small sample size could not provide definitive estimates of interjudge reliability at that time. Item agreement appeared adequate because judgments fell within 1 point or less on a 5-point scale for nearly all items.

The instructor pilots independently watched the remaining 46 videotapes of the flight scenarios in the high-fidelity simulator. For each team, the judges (a) listened to the audiotaped brief, (b) viewed the video of the pilots in the simulator, and (c) completed the problem- and process-oriented rating forms.

Scoring the Explicit Measurements From the Low-Fidelity Trainer Scenario

Two raters who were not familiar with the performance of the crews in the full-mission simulator scenario independently scored the pilots’ responses. A scoring guide was developed to aid in the reliable evaluation of each participant’s response to the questions he or she was given in the low-fidelity trainer scenario. The guide contained a 5-point behaviorally anchored rating scale. The form listed possible pilot responses and their corresponding point values, based on the closeness to the correct response (response accuracy and point values were previously determined by an experienced flight instructor). Responses were assigned 5 points for the optimal answer, and 2, 3, or 4 points based on the closeness to the correct answer. For example, one question asked, “What are your current cloud tops (the altitude of the top of the clouds)?” The answers were scored as follows: 5 = 1,000 ft; 4 = 900 to 999 or 1,001 to 1,100 ft; 3 = 800 to 899 or 1,101 to 1,200 ft; 2 = any other answer than listed for 3, 4, or 5 points; and 1 = no answer given.

Although the questions were designed to maximize objectivity, a few required a response phrase for the answer. One such question was “What was the last ATC call?” The optimal answer had three components and each had to be evaluated by the scorers.

For the crew members, the flight was a single flight; however, for data analysis, the three question sets collected at the three stopping points were analyzed both separately and together. Team situation awareness was measured by adding the scores of pilots and copilots to represent the total flight knowledge possessed by the crew.

Simple addition of the individual members’ scores to obtain a team score is clearly not the only approach that might be used. The implications of one, both, or neither crew member having the relevant flight knowledge may be quite different. Varying task demands for the pilot flying and the pilot not flying may also influence the importance of who has what knowledge in a particular context. However, a simple additive approach can be justified in this simulation because both crew
members had access to the same visuals and the same ATC communications and each had almost equal need for the information. In fact, several crews switched roles for some portion of the flight. Each crew member had an assigned task (one to fly, the other to communicate) that is not demanding in this scenario. Whether the validity could be improved through means other than a simple combination of scores deserves further research.

ANALYSES AND RESULTS

Explicit Assessment Reliability

Both raters scored each of the 48 responses for both pilots and copilots. The mean point value assigned by the two raters was used for a participant’s score. Agreement between the judges was evaluated using Pearson correlations. Correlations were computed between the raters for each of the 48 questions asked. Rater agreement for the correctness of the responses was significant \((p < .01)\). The mean correlation was .88, and the median was .92 (range = .52 to 1.00). Illegible responses were not evaluated.

Internal consistency for responses was examined both within and across the three question sets. Scores for teams were also examined for consistency. Three estimates of internal consistency were computed. First, all item responses across all questions were analyzed to assess the internal consistency of the measurement instrument as a whole. Second, item responses within each question set were analyzed to assess the internal consistency of each. Third, scores were computed within sets and correlated across sets. For the within-sets scores, pilots and copilots received as their score the mean of the responses within that question set; this was done to maximize the available sample size for analyses of scores. Team scores were also computed by adding the pilot and copilot scores for that item.

Estimates of internal consistency (Cronbach’s alpha) can be found in Table 1 along with sample sizes and numbers of items for pilot, copilot, and team scores. As can be seen in Table 1, the estimates are acceptable or better with the exception of Question Set 2 for the pilots. Correlations and descriptive statistics for the three question sets are shown in Table 2. Although the scores for pilots showed somewhat smaller correlations among sets than the scores of copilots, tests for the differences of the correlations between pilot and copilot correlations were not significant \((z\) for Segment 1 vs. Segment 2 was \(-.69\); \(z\) for Segment 1 vs. Segment 3 was \(-.24\); and \(z\) for Segment 2 vs. Segment 3 was \(-1.77\), all showing \(p > .05\)).

Implicit Measurement Reliability

The reliability of the implicit ratings was assessed by computing alpha with judges treated as items. For the problem ratings, the reliability estimates were
quite satisfactory, with a mean estimate of .78 and a median of .81. For the team situation awareness ratings, the reliability estimate (alpha for judges as items) was .87.

Effects of Repeated Testing

If participants are cued by exposure to questions tapping flight-related knowledge and their attention is directed toward the question elements, then scores should increase after the first question set. To assess whether testing appeared to influence test scores, the scores for the 11 repeated items were computed across all question sets. For each test item, a repeated measures analysis of variance was computed to determine whether significant differences across repetitions occurred. Results of the tests can be found in Table 3. Two items (Items 3 and 5) were repeated only once.

As can be seen in Table 3, 9 of 11 $F$ tests were significant. However, for only 4 of the items did the means increase when they were repeated. One such item con-

---

**TABLE 1**

Internal Consistency Within Each Segment by Crew Position

<table>
<thead>
<tr>
<th></th>
<th>Segment 1</th>
<th></th>
<th>Segment 2</th>
<th></th>
<th>Segment 3</th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha$</td>
<td>$N$</td>
<td>$\alpha$</td>
<td>$N$</td>
<td>$\alpha$</td>
<td>$N$</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>Pilot</td>
<td>.62</td>
<td>38</td>
<td>.02</td>
<td>37</td>
<td>.69</td>
<td>38</td>
<td>.66</td>
</tr>
<tr>
<td>Copilot</td>
<td>.59</td>
<td>37</td>
<td>.70</td>
<td>37</td>
<td>.71</td>
<td>36</td>
<td>.84</td>
</tr>
<tr>
<td>Team</td>
<td>.71</td>
<td>34</td>
<td>.59</td>
<td>34</td>
<td>.76</td>
<td>33</td>
<td>.77</td>
</tr>
</tbody>
</table>

*Note.* Differences in $N$ reflect items that were not scored due to illegibility.

18 items. b14 items. c16 items. d48 items.

**TABLE 2**

Correlations Among Segments and Descriptive Statistics

<table>
<thead>
<tr>
<th>Segment</th>
<th>Pilot</th>
<th></th>
<th>Copilot</th>
<th></th>
<th>Team</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1.00</td>
<td>—</td>
<td>—</td>
<td>1.00</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>0.28</td>
<td>1.00</td>
<td>—</td>
<td>0.55**</td>
<td>1.00</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>0.17</td>
<td>0.25</td>
<td>1.00</td>
<td>0.32*</td>
<td>0.58**</td>
<td>1.00</td>
</tr>
<tr>
<td>$M$</td>
<td>2.89</td>
<td>3.21</td>
<td>2.82</td>
<td>2.91</td>
<td>3.14</td>
<td>2.90</td>
</tr>
<tr>
<td>$SD$</td>
<td>0.53</td>
<td>0.38</td>
<td>0.54</td>
<td>0.53</td>
<td>0.60</td>
<td>0.53</td>
</tr>
</tbody>
</table>

*Note. N = 41.*

*p < .05. **p < .01.
concerned the controlling agency (Who is your current ATC?) and the other 3 all concerned traffic. For the other 5 significant tests, means decreased over repetitions.

Validity

Pearson correlations were calculated to provide an estimate of the relation between overall team situation awareness scores in the high-fidelity scenario and performance across all emergencies and overall team performance (see Table 4). Team situation awareness as measured by the implicit method in the high-fidelity simulator was significantly correlated with performance across all emergencies in that simulation and with overall team performance in that simulation.

Because performance measures were not appropriate for the scenario on the low-fidelity trainer, we predicted that differences in factual knowledge as measured in that scenario would be associated with instructor ratings of performance on the flight flown in the full-motion simulator. The items from the problem-oriented rating form were used for this purpose. Pearson correlations were calculated to provide an estimate of the relation between team scores from the questions in the low-fidelity scenario and each of the seven problem-oriented ratings, individually and overall (see Table 5). Significant correlations were found between team ratings on six of the seven problem-oriented outcomes and the scores earned in the low-fidelity trainer scenario. Given the reliability estimate for both the ratings based on question responses and the instructor ratings, the observed correlations indicate substantial relations between the assessment and both types of ratings at the true score level.

---

### Table 3

<table>
<thead>
<tr>
<th>Item</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Current heading</td>
<td>9.56</td>
<td>8.11</td>
<td>7.41</td>
<td>2/77</td>
<td>22.86*</td>
</tr>
<tr>
<td>3. Controlling agency</td>
<td>7.01</td>
<td>8.57</td>
<td>—</td>
<td>1/40</td>
<td>11.89*</td>
</tr>
<tr>
<td>4. Last air traffic control call</td>
<td>6.26</td>
<td>5.67</td>
<td>5.29</td>
<td>2/79</td>
<td>2.24</td>
</tr>
<tr>
<td>5. Weather at destination</td>
<td>4.02</td>
<td>4.28</td>
<td>—</td>
<td>1/39</td>
<td>0.45</td>
</tr>
<tr>
<td>6. What airport would you use right now for an emergency</td>
<td>9.01</td>
<td>5.48</td>
<td>4.74</td>
<td>2/78</td>
<td>65.64*</td>
</tr>
<tr>
<td>7. Airport bearing</td>
<td>6.64</td>
<td>4.44</td>
<td>4.36</td>
<td>2/67</td>
<td>18.03*</td>
</tr>
<tr>
<td>8. Airport range</td>
<td>5.84</td>
<td>4.49</td>
<td>4.25</td>
<td>2/67</td>
<td>7.69*</td>
</tr>
<tr>
<td>9. Distance to traffic</td>
<td>3.51</td>
<td>3.56</td>
<td>5.98</td>
<td>2/79</td>
<td>24.16*</td>
</tr>
<tr>
<td>10. Clock position of traffic</td>
<td>3.78</td>
<td>5.15</td>
<td>7.56</td>
<td>2/80</td>
<td>31.10*</td>
</tr>
<tr>
<td>11. Altitude of traffic</td>
<td>4.24</td>
<td>4.58</td>
<td>7.89</td>
<td>2/79</td>
<td>27.07*</td>
</tr>
</tbody>
</table>

*p < .01.
The correlation between the team situation awareness scores from both measurement methods was significant ($r = .43$, $p < .01$). The correlation of the two measurements for the copilot was also significant ($r = .46$, $p < .05$), but the correlation between the two scores for the pilot alone was not significant ($r = .25$).

**DISCUSSION**

The results from this research are encouraging for those interested in advancing knowledge about situation awareness. We have demonstrated that team situation awareness can be reliably measured with two different measurement methods, each of which is related to performance. This can occur even when team situation awareness and performance are measured in entirely different scenarios.
is promising for theory development and future research that the two different measures of team situation awareness are significantly correlated. This suggests that despite differences in focus (action vs. cognition) they appear to measure at least in part the same underlying construct.

Although the correlation between the two measures of team situation awareness was significant, it was low. Explanations for this include the equipment used, the task demands in the scenarios, and the different focus of the two measurement methods. The training devices (i.e., high-fidelity simulator and low-fidelity desktop trainer) used in the scenarios differed in the cues they could provide for awareness. The two scenarios used made different awareness demands on the crews. Situation awareness in one scenario was measured primarily on the knowledge of elements outside the cockpit (weather, traffic) and in the other simulation, awareness was measured primarily when crews reacted to equipment problems. This difference could be important at the experience level of the crew members (Prince & Salas, 1998).

The nonsignificant correlation between the situation awareness of the team, as measured by the explicit measurement method and performance, on the first emergency is illustrative. This emergency occurs just after take-off, when the pilot flying is engaged in the technical aspects of flying and the copilot is busy completing required checklists. The technical flying task for the pilot is particularly demanding at this time. He or she is involved in the tasks of taking off, ascending to the assigned altitude and getting on his or her flight path and also, in the case of the simulator, is trying to adjust to equipment that is demanding to operate. (The simulator, according to instructors and students both, is more difficult to fly than the actual aircraft.) As the icing progresses, the aircraft becomes ever more difficult to control. The instrument that gives the best information about icing is located in the ceiling of the cockpit, over the pilots’ heads. Checking this instrument would require the pilot flying to interrupt his or her scan and attention to the flying task. The pilot flying is applying technical knowledge and skill in basic aircraft flight and his or her awareness is focused there. Situation awareness, as measured by observation in the scenario, was considered in the light of the tasks, but situation awareness, as an overall measurement of ongoing flight knowledge in a routine flight, is not an important part of the pilot’s task at that moment.

This also suggests that workload does affect basic flight knowledge of aviators at a low level of flight experience even though a general independence between situation awareness and workload has been demonstrated for other populations (Endsley, 1993). Pilots at this level must still focus more of their attention on basic flying tasks than do those whose experience has allowed them to execute some tasks more or less automatically.

Perhaps the biggest difference in the measurements is the different bases for the two measurement methods (i.e., cognitive for one, cognitive expressed in action in the other). Situation awareness, as measured by its cognitive components, may not
result in appropriate action for various reasons. For an individual, a failure to act may be due to ignorance of the appropriate action. For a team, additional reasons for inaction include inhibition of action due to status within the team, lack of control over the necessary action, an expectation that another will take the action, and actions taken by the other team member that preclude one’s own actions.

Implications for Training

First, it is now possible to evaluate the effects of training in team situation awareness because reliable measurement of team situation awareness is now feasible. Next, we confirmed a need for training in team situation awareness in this population. Scores on the explicit measurement showed a wide range. This is a concern because these scores were based on answers to questions that were all considered by senior pilots to be important to situation awareness within the simulation.

There is some consistency to team situation awareness; crews with higher ratings in one scenario tended to be those with higher ratings in another. Although one’s awareness may change from moment to moment during a flight, the team’s ability to gain and maintain awareness appears to be relatively stable within a period of time. When coupled with the apparent effect of experience on situation awareness (Carretta et al., 1996; Prince & Salas, 1998), this consistency has important implications for training. It suggests that situation awareness can be developed and that progress in its development can be reliably measured. These are two prime requirements for training.

Based on this research, we recommend the use of low-fidelity trainers for some aspect of the training for low experience level aviators. We have demonstrated that there is a relation between situation awareness in simulations on the low-fidelity trainer and on the simulator that more closely resembles real flight. Aviators with high experience levels receive practice and feedback on situation awareness in a realistic scenario during their annual recurrent training. This is generally considered sufficient. However, new aviators, lacking the experience that helps build the ability to gain and maintain situation awareness, need more intensive training. It is possible that simulations of routine flight on a low-fidelity trainer, combined with an explicit measure of situation awareness, may be effective for training team situation awareness. Simulations on a low-fidelity trainer are easier to develop and implement and instructors can interrupt the scenarios to question and give immediate feedback on the crew’s flight knowledge as they progress through the scenario. We recommend that research into design, implementation, and effectiveness of a training program for aviation team situation awareness aimed at the low experience level aviator be done.

Changes in understanding of situation awareness and its requirements develops over many hours in the cockpit (Prince & Salas, 1998). For this reason, we encour-
age further study of situation awareness with professional aviators of varying experience levels when they are carrying out real tasks.

ACKNOWLEDGMENTS

This research and development reported here was also partially supported by the Federal Aviation Administration Human Factors Research and Engineering Group (ATOP–Human Factors; Terry Allard, Program Director). All opinions stated in this article are those of the authors and do not necessarily represent the opinion or position of the University of Central Florida or the Federal Aviation Administration.

REFERENCES


Manuscript First Received: December 2005