

## The Utility of Event-Based Knowledge Elicitation

Jennifer E. Fowlkes and Eduardo Salas, University of Central Florida, Orlando, Florida, David P. Baker, American Institutes for Research, Washington, D. C., Janis A. Cannon-Bowers, Naval Air Warfare Center Training Systems Division, Orlando, Florida, and Renée J. Stout, Alignmark, Orlando, Florida.

The purpose of this investigation was to describe and evaluate an event-based knowledge elicitation technique. With this approach experts are provided with deliberate and controlled job situations, allowing investigation of specific task aspects and the comparison of expert responses. For this effort a videotape was developed showing an instructor pilot and student conducting a training mission. Various job situations were depicted in the video to gather information pertinent to understanding team situational awareness. The videotape was shown to 10 instructors and 10 student aviators in the community, and responses to the videotape were collected using a questionnaire at predetermined stop points. Consistent with expectations, the results showed that more experienced respondents (i.e., instructors) identified a richer database of cues and were more likely than students to identify strategies for responding to the situations depicted, providing some empirical evidence for the validity of the event-based technique. This method may serve as a useful knowledge elicitation technique, especially in the later stages of a job analysis when focused information is sought.

### INTRODUCTION

Knowledge elicitation is a component of knowledge acquisition in which information pertaining to the reasoning and other thought processes needed to perform a job is obtained from a human source. Knowledge elicitation has become an increasingly important task in modern work environments where understanding the cognitive requirements associated with highly complex jobs is critical. It is also a task associated with frustration (e.g., prying information from experts), large time investments (e.g., in coding, collecting, and analyzing verbal protocols), and, worst of all, "art," in that the quality of the information received depends on the interviewer's technique and experience (Cooke, 1994; Duda & Shortliffe, 1983; Hoffman, 1987).

Considering the importance of knowledge elicitation as well as its inherent difficulties, it is necessary that the analyst be equipped with a variety of techniques that can be selected based on their suitability to the problem at hand. The purpose of this paper is to describe an assessment of one such technique, referred to as *event-based knowledge elicitation*. *Event-based* means that the expert is provided with known and controlled job situations. These are selected because prior analysis (e.g., interviews with subject matter experts) suggests that experts' reaction to them will reveal meaningful information about specific aspects of the job.

This approach is similar to case study (Cooke, 1994) and test case (Hoffman, Shadbolt, Burton, & Klein, 1995) protocols. They allow a priori expectations to be developed and allow data to be collected from a number of

experts on the same stimulus set so that their responses can be compared. Although trade-offs occur in the application of any knowledge elicitation technique, these approaches appear to yield useful information when focused information is sought. However, apart from their apparent usefulness, little is known empirically about the validity of such approaches. As Hoffman et al. (1995) noted, it is not enough to know that a particular technique appears useful; other types of evidence are needed.

In order to provide some context for event-based knowledge elicitation, Table 1 provides examples of "direct" knowledge elicitation approaches – that is, approaches that obtain knowledge by directly asking or observing the expert. Several general comments can be made

about these techniques. First, in one way or another, all are situated or embedded in a job context. In a structured interview, for example, an expert may be prompted to recall a job situation. In the method of familiar tasks (Hoffman, 1987), experts are instructed to provide a commentary on their actions as they perform their job. The embedded aspect of knowledge elicitation is critical for revealing meaningful information about the job (Hoffman, 1987).

A second observation is that no single method can provide a complete characterization of the cognitive aspects of task performance. Different approaches are better suited to different stages of a job analysis, a process that is generally iterative. Techniques such as

**TABLE 1:** Examples of Knowledge Elicitation Techniques

Knowledge Elicitation Method	Example Application	Characteristics
Unstructured interview	Free-form discussions with an expert.	Entails little preparation but may require extensive domain knowledge on part of analyst to gain useful information. Provides little control over task content. Most applicable early in job analysis. Coding and analysis of data problematic.
Structured interview	Discussions with an expert in which the topics covered and sequencing are largely predetermined.	May be useful at any time during a job analysis. Entails more preparation than an unstructured interview and may require extensive domain knowledge on the part of the analyst to gain useful information. Control over the specific aspects of the job treated varies with the application.
Familiar tasks	Analyst seeks to obtain, via direct observation or verbal protocols, a sample of experts performing typical tasks.	Provides a foundation for job analyses such as task analysis.
Tough cases (Hoffman, 1987)	Analyst seeks to obtain, via direct observation or verbal protocols, examples of experts in challenging job situations.	Most applicable late in job analysis, when focused information is sought.

the unstructured interview are more applicable to the front end of the analysis when the analyst is learning about the domain and identifying relevant variables. Other techniques, such as the method of tough cases, have greater utility when focused information is sought, after a knowledge database is well established (Hoffman, 1987; Hoffman et al., 1995). It has also been argued via the differential access hypothesis that different elicitation techniques are better suited to eliciting different types of knowledge (Hoffman et al., 1995).

A final observation, and one directly pertinent to the present paper, is that techniques vary in their control over the job content treated during the knowledge elicitation. At one extreme is the unstructured interview. By definition, in the unstructured interview the expert may discuss aspects of the job that cannot be anticipated beforehand. In the method of familiar tasks, the goal for task constraint is to observe the expert performing a sample of representative tasks. In the method of tough cases (Hoffman, 1987), generally job situations are sought that are challenging to the expert to reveal important facets of the expert's reasoning or problem-solving approaches. These job situations may not be known beforehand. Indeed, a disadvantage of this approach is that as data are collected in the actual job environment, tough cases occur unpredictably. In some respects data obtained for knowledge engineering are left to chance.

Event-based knowledge elicitation differs from other approaches mainly in its emphasis on exerting control over the stimulus presentation to the expert. Such a method can be applied during the later stages of a job analysis, when the analyst seeks to address specific aspects of the task. For example, in the air traffic control domain, it is known that experts categorize aircraft into event types to minimize workload and enable rapid retrieval (Redding, Cannon, & Seamster, 1992). Thus to learn more about the categorization schemes used by experienced controllers, the analyst could develop scenarios in which different traffic event types were presented. Through simulation, these same event types could be shown to a number of controllers and their responses compared.

Event-based techniques are finding their way into training and performance measurement realms (Fowlkes, Dwyer, Oser, & Salas, 1998; Fowlkes, Lane, Salas, Franz, & Oser, 1994; Johnston, Smith-Jentsch, & Cannon-Bowers, 1997). In essence these techniques seek to make explicit links among (a) goals, (b) scenario or exercise design, and (c) resulting output. Table 2 compares the event-based methodology in training and knowledge elicitation contexts. In a training context, events are included in an exercise to provide known opportunities for trainees to perform tasks targeted in the training (i.e., that are specified by the training objectives). This method ensures that training opportunities are not left to chance. Because events are known beforehand, expectations for trainee responses can be developed and incorporated into performance measurement. Assessments of trainee performance are tied to how they respond to the events. The benefits include standardization, reduction of workload for the instructor (e.g., not everything has to be observed), and diagnostic performance measurement.

When event-based knowledge elicitation is implemented, similar linkages are established and similar benefits may result. The linkages include (a) establishing knowledge elicitation goals and hypotheses, (b) developing a scenario that includes events that prompt experts to provide information in the key areas, and (c) development of a priori expectations and analysis techniques. The benefits include economy of effort (not everything the expert says or does has to be captured – only those things that pertain to the targeted tasks/events) and the ability to capture a large number of expert responses to the same, highly meaningful tasks.

Although variants of event-based knowledge elicitation are employed by analysts, little is known about their validity. If such techniques were valid for the elicitation of data from experts, we would expect to see differences between respondents based on their experience. The purpose of this investigation was to examine the utility of the event-based knowledge elicitation technique for obtaining information pertinent for team situation awareness (SA) in a military helicopter community. To evaluate this method, we used a

TABLE 2: Comparison of Event-Based Approaches

Comparisons	Event-Based Application	
	Training	Knowledge Elicitation
Links	<p><i>Goal.</i> Training objectives.</p> <p><i>Exercise design.</i> Events included to provide known opportunities for trainees to respond.</p> <p><i>Output.</i> Development of performance measures and data collection tailored to the events and that provide diagnostic performance assessments in relation to training objectives.</p>	<p><i>Goal.</i> Information sought about a particular subdomain or reasoning strategy.</p> <p><i>Exercise design.</i> Events designed to provide data directly relevant to the knowledge elicitation goals.</p> <p><i>Output.</i> A priori data coding and analyses can be developed to address knowledge elicitation goals.</p>
Benefits	<p><i>Economy.</i> Not all trainee actions have to be observed, only those that are reactions to scenario events.</p> <p><i>Data capture.</i> Instructors can position themselves to observe trainee reactions to events so that key behaviors are not missed. Because events are known a priori, measurement tools can be tailored.</p> <p><i>Diagnosticity.</i> Allows focused, meaningful feedback to be provided.</p> <p><i>Standardization.</i> All trainees can receive training in key areas.</p>	<p><i>Economy.</i> Constrains data collection to expert's reactions to the exercise events.</p> <p><i>Data capture.</i> Analysts can position themselves to observe trainee reactions to events so that data collection opportunities are not lost.</p> <p><i>Diagnosticity.</i> Because events are known a priori, data collection tools, coding procedures, and databases can be tailored before the data collection events.</p> <p><i>Standardization.</i> All experts can react to the same events, allowing their output to be compared.</p>

variation in which a videotape was produced that showed an instructor pilot and student conducting a training mission in which various job situations were deliberately built in. We developed the events in the videotape to target data theorized to be important to understanding team SA (Stout, Cannon-Bowers, & Salas, 1994). These are (a) the cues and patterns in the situation assessed by team members, including actions of other team members, and (b) the nature and content of shared mental models that allow team members to interpret and react to cue information from the environment.

The videotape was shown to 10 instructor pilots and 10 student aviators in a military helicopter community. We predicted that more experienced respondents would identify a

richer database of cues and would be more likely than students to identify strategies for responding to the situations depicted in the videotape.

## METHOD

### Participants

The participants were 20 military helicopter pilots, of whom 10 were instructors in the training command and 10 were newly winged student aviators. The instructors (one woman and nine men) possessed an average of 253 h ( $SD = 98.78$ ) in the training aircraft and 1361 flight h overall ( $SD = 602.82$ ). Pilots in the student group (two women and eight men) possessed an average of 115 flight h ( $SD =$

13.52) in the training aircraft and 190.50 h overall ( $SD = 52.50$ ).

### Materials

**Videotape.** The videotape depicted an instructor under training (IUT) and an instructor pilot (IP) performing a night training flight from the preflight brief through the flight. The videotape included specific cue information pertaining to team interactions and flight situations to elicit the experts' reactions. Stops in the videotape provided standardized points at which the pilots' responses to questionnaire items about the segment just viewed would be obtained. Each segment of the videotape is summarized in Table 3.

The content of the videotape was determined through structured interviews with experienced aviators regarding the cues important for maintaining team SA. When filming, the pilot actors adhered to a script that detailed

the scenario events. During the initial segment, the pilots were depicted sitting at a table conducting a preflight brief. The remainder of the training flight was videotaped in a full mission flight simulator. This portion was filmed with a single VHS camera located behind the pilots. The shot depicted the flight instruments and system gauges. All communications that occurred were recorded as well.

**Questionnaire.** A questionnaire was developed to obtain information about the cues and team processes used to acquire and maintain team SA. Items included a warm-up question in which participants provided an overview of the segment just viewed. A second item instructed respondents to identify the pertinent cues for maintaining team SA. These could include instruments and gauges, external cues, communications, and procedures. Responses to this item were coded using the cue codes described later in this paper. There was 90%

TABLE 3: Overview of Videotape

Segment	Description
1: Preflight brief	Critical cues built into the preflight brief portion of the videotape included a night flight, marginal weather, and low outside air temperature, which, combined with the planned flight altitude, should signal potential icing conditions. In addition, the aircraft had a previous history of problems in systems critical for the flight. Finally, poor briefing habits were displayed, including lack of eye contact and cursory treatment of emergency procedures.
2: Takeoff to level off	This segment covered the performance of preflight checks through to the climbout. Cues included lack of verbalization of the checklists and crew member actions, and communications with ATC that were delayed. During the takeoff portion of the flight, cues included change of aircraft controls, being cleared to a higher altitude than expected because of traffic in the area, and setting up for the approach. At the end of the segment the airspeed decreased.
3: Icing	The crew diagnosed a drop in airspeed as caused by icing, performed applicable icing checklists, and requested a descent from air traffic control. Predictable, secondary signs began to occur, such as a fluctuating loadmeter, because of the drain on the electrical system caused by the icing systems. Other cues included little verbalization of actions on the part of the crew members and little forward planning.
4: Generator failure	A generator failure occurred and the generator had to be reset twice. The crew also prepared for an instrument approach. They discussed whether to declare an emergency but made no decision and did not inform air traffic control of the flight status. Cues pertaining to crew interaction included assertiveness on the part of the junior pilot, not verbalizing the performance of checklists, and lack of decision making.

agreement between the two coders used to code this questionnaire item. Discrepancies were handled through consensus.

The third questionnaire item instructed participants to identify information that the crew should be considering or sharing. Responses to this item were coded into the shared mental models categories to be described later. There was 85% agreement between the coders in coding the responses to this item, and discrepancies were handled through consensus.

A third questionnaire item required that participants make predictions for the next flight segment. Responses to this item were not analyzed because of the generally poor responses that were obtained.

### Procedure

Pilots were tested in groups of two in sessions that lasted approximately 1.5 h. Each session began with an overview of the purpose of the study followed by a description of the flight (background information) and weather depicted in the videotape. Participants were provided approach plates, a chart, a pocket checklist, and a cockpit diagram, to which they could refer at any time. In addition, paper was provided for recording notes, which they were free to make at any time. The task was to watch the videotape of the helicopter flight from the initial briefing through the flight. At each of the four planned stops, participants individually responded to the questionnaire. Prior to each flight segment, an overview of the flight situation was provided (e.g., "This segment picks up where the last segment left off.").

### Coding Taxonomies

Coding taxonomies were developed and employed to allow comparison of results across the two groups on issues directly relevant to team SA. The taxonomies were developed in a pilot effort using the videotape just described in combination with a structured interview. Interviews were conducted with 12 instructor pilots from the community (not the instructors used for the present study). These data were combined with another set of data from another military aviation community in

which a similar technique was used but in which a paper-based scenario was presented to respondents.

Development of the taxonomies involved a two-step procedure. First, the responses to each question were listed for each participant. Second, results from the data lists for each interviewee were reviewed and summarized into a master list of cues and shared mental models. In total, two master lists were developed – one for each data set. Duplicate cues and shared mental models were combined or eliminated during this process. Next, the two master lists were combined into a single list.

Cues and model categories and their corresponding codes were developed from the combined master list. Categories were structured to reflect theories of situation assessment (Endsley, 1995; Salas, Prince, Baker, & Shrestha, 1995; Stout et al., 1994), teamwork (e.g., see Salas, Dickinson, Converse, & Tannenbaum, 1992), and shared mental models (Converse & Kahler, 1992; Rasmussen, 1986). Each category is briefly described in the following paragraphs.

*Cue category codes.* Cues were defined as any information associated with the task and environment that is processed by an individual to develop an understanding of a situation. In total, three categories of cues were identified from the master list: task-based cues, team-based cues, and environmentally based cues. Each consisted of more specific variables, which were coded. From the master list, nine codes for task-based cues were identified: instruments, navigation, checklists, charts/maps/approach plates, control of aircraft, checkpoint, wing position (not applicable to the present scenario), radar, and aircraft sounds/vibrations. Team-based cues were those associated with other team member behaviors and communications, as well as communications from other supporting air traffic agencies. From the master list, a total of three team-based cue codes were identified: intracrew communications, crew member behaviors, and external communications. Finally, environmentally based cues were those associated with the environment external to the aircraft cockpit. From the master list, a total of five environmentally based cue codes were identified:

weather, external aircraft condition, time of day, obstacles, and terrain.

*Shared mental model category codes.* Shared mental models were defined as organized bodies of knowledge that are shared across team members (Cannon-Bowers, Salas, & Converse, 1993). A review of the literature suggested that three types of mental models could be shared among team members that were important for situation assessment: declarative models, procedural models, and strategic models (Converse & Kahler, 1992).

Each type of mental model served as a global coding category for which more specific codes were defined from the master data list. First, declarative models consisted of main concepts, facts, and rules associated with the missions. Codes under declarative models included knowledge of roles and responsibilities, cockpit configuration, weather, condition of aircraft, crew experience levels, mission goals, aircraft systems, obstacles, terrain, checkpoints, and publications. Second, procedural models consisted of knowledge associated with the sequence and timing of activities required to complete tasks in each mission. Codes under procedural models included knowledge of crew member action/tasks, standard/emergency procedures, timing for mission, aircraft position during mission, and

wing position during mission (not applicable to the present scenario). Last, strategic models consisted of strategies that allow team members to apply their declarative and procedural knowledge to specific task situations. Codes under strategic models consisted of knowledge and strategies associated with current plan, current state, status, and deviations.

## RESULTS

### Cues Identified

Figure 1 displays the average number of cues identified for the instructor and student groups across the four videotape segments. It can be seen that at each segment, instructors identified more cues than did students. In addition, the number of cues identified by both groups increased across segments, possibly because of a practice effect. These observations were confirmed with a mixed-model ANOVA, with experience (instructor versus student) as the between-subjects factor and segment as the repeated factor. The results revealed an effect caused by experience,  $F(1, 18) = 7.36, p < .05$ , and segment,  $F(3, 54) = 7.77, p < .05$ . The Experience  $\times$  Segment interaction was not significant,  $F(3, 54) = 1.44, p > .05$ .

Figure 2 shows the frequencies for each of the three cue types across the four segments.

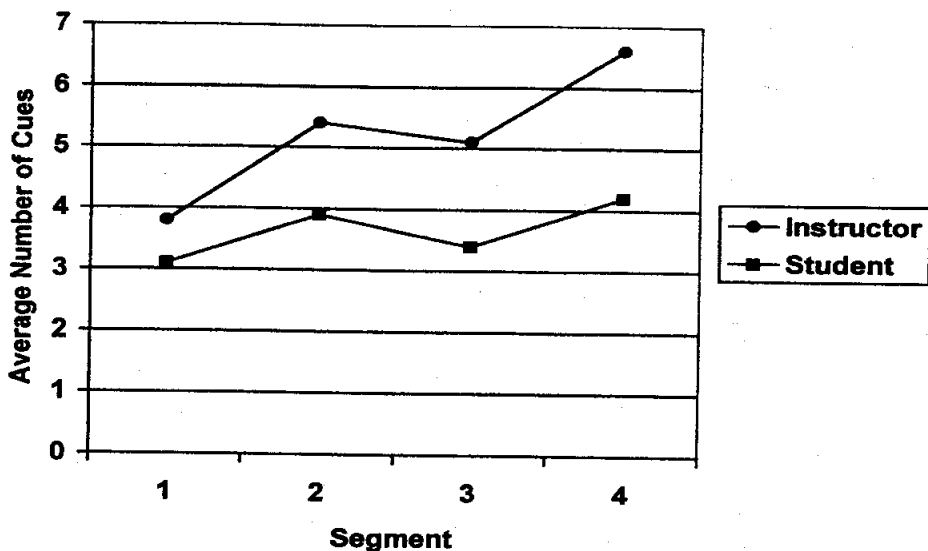


Figure 1. Average number of cues identified by instructors and students across segments.

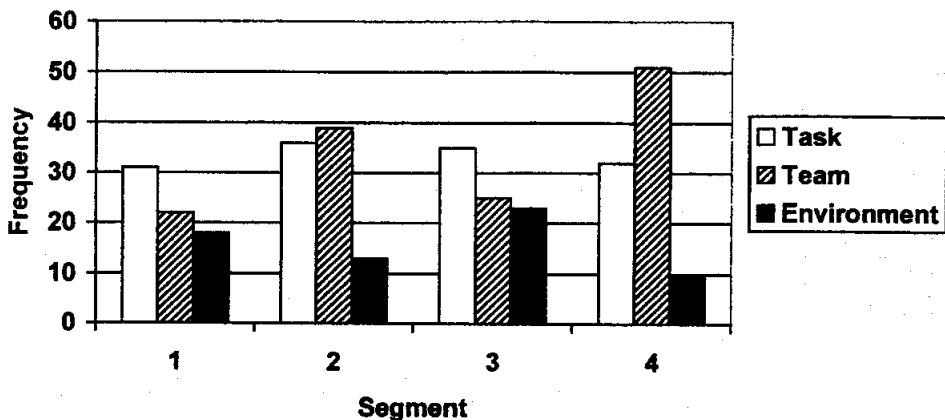


Figure 2. Frequencies of responses coded into task, team, and environmental cue categories for each segment.

These data combine the responses from instructors and students because a chi-square analysis revealed no differences between groups in terms of the proportion of cue types that were listed as being important for developing and maintaining team SA,  $\chi^2(2) = 0.33$ ,  $p > .05$ . As shown in Figure 2, no trends across segments are evident for task and environmental cues. However, there appear to be more team-related cues identified across the segments, and team-based cues were especially prominent in Segment 4. This may account for the overall increase in cues across segments seen in Figure 1. Whether this is caused by a practice effect or by the nature of Segment 4 is uncertain.

To provide additional information on cues identified as being important for maintaining team SA, Table 4 summarizes the most prominent cues for the task, team, and environment categories – that is, cues that were identified by at least four participants (20% of the sample) are shown. During the flight (Segments 2–4), task-based cues identified consistently across the segments included the use of checklists and the quality of basic air work. In addition, a variety of team cues were identified by respondents consistently across segments. These included backup behavior, verbalizing upcoming actions before they were performed (e.g., such as changing radios or navigation aids), and planning actions.

Finally, prominent environment cues included time of day, general weather conditions,

outside air temperature, and icing. Weather was the only environmental cue consistently identified across all segments.

### Shared Mental Models

Figures 3 and 4 show for instructors and students, respectively, the number of responses categorized in each model category. It can be seen that the strategic model category was the most prominent for instructors in at least three of the four segments. That is, when asked what the crew should be communicating, instructors predominantly made responses pertaining to strategizing and planning. These responses included terminating the flight, descending to achieve visual flight rules (VFR) conditions, preparing for a possible generator failure, and, once a generator failure occurred, determining what systems could be cut off to conserve electrical power.

For students (Figure 4), discussions of the procedural aspects of the task appeared to be most common. These responses included improving basic air work, completing checklists, and following emergency procedures. The apparent differences between the instructors and students was confirmed by a chi-square analysis,  $\chi^2(2) = 10.17$ ,  $p < .05$ .

### DISCUSSION

Variations of event-based knowledge elicitation are routinely used to support job analyses. The purpose of this study was to examine the



effect of experience on responses obtained using this approach. In terms of the identification of relevant cues for maintaining team SA, we found no differences between instructors and students in the types of cues that were identified, but we did find that instructors identified significantly more cues, in accordance with expectations. There was also a trend toward the identification of more cues across segments for both students and instructors, possibly because of a practice effect. Finally, when asked what type of information the crew should be sharing, the instructors and students differed in the type of information identified. Students were more likely to identify procedural, "how-to" aspects of the task, such as completing checklists, whereas instructors were more concerned that the crew should be developing or discussing strategies – that is, they focused on identifying what should be done, consistent with expert-novice differences reported in the literature (Ericsson & Lehmann, 1996).

The results were obtained using a questionnaire to collect responses, a format that might

be expected to limit the richness and breadth of the responses. However, such a format has advantages, such as ease of coding and analysis, a significant benefit given the laborious efforts expended during the data coding and analysis phase characteristic of many knowledge elicitation techniques.

To place the advantages and drawbacks of this technique in fuller perspective, we will now critique it based on criteria identified by Hoffman (1987).

### Simplicity

As defined by Hoffman (1987), *simplicity* refers to how easily the task is understood by participants and the nature of the materials that must be developed to support the task. We found that the task was easily understood by participants and straightforward to implement. The one problem noted was that there was hesitation on the part of some of the participants to pass judgment on the performance of another aircrew member. Thus we had to emphasize in the instructions that the aviators were acting according to a script.

**TABLE 4:** The Most Common Task, Team, and Environmental Cues Extracted across Segments

Cue Category	Segment			
	1	2	3	4
<b>Task</b>				
Aircraft condition	✓			
Emergency procedures	✓			
Checklists		✓	✓	✓
Control of aircraft		✓		
Instruments			✓	✓
Basic airwork	✓	✓	✓	✓
<b>Team</b>				
Crew communications				
Verbalize actions		✓	✓	✓
Discuss/verbalize plans	✓		✓	✓
ATC communications		✓		✓
Crew behavior				
Attitude/assertiveness	✓			
Backup		✓	✓	✓
Synchronize actions	✓	✓	✓	✓
<b>Environment</b>				
Weather	✓	✓	✓	✓
Time of day	✓		✓	
External aircraft condition			✓	

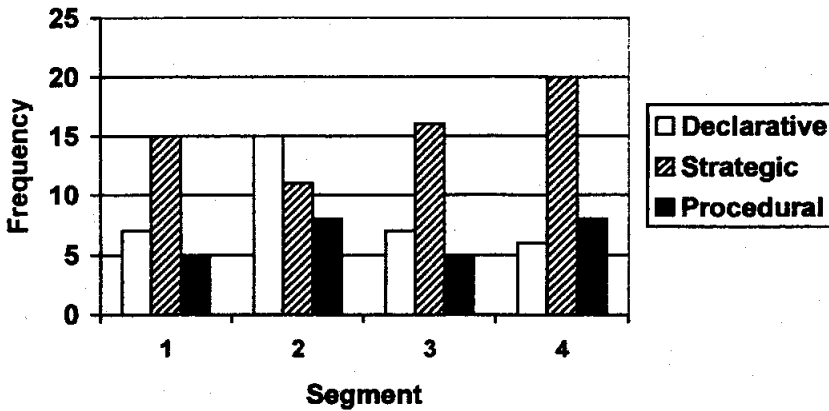


Figure 3. Frequencies of instructor responses coded into each mental model category for each segment.

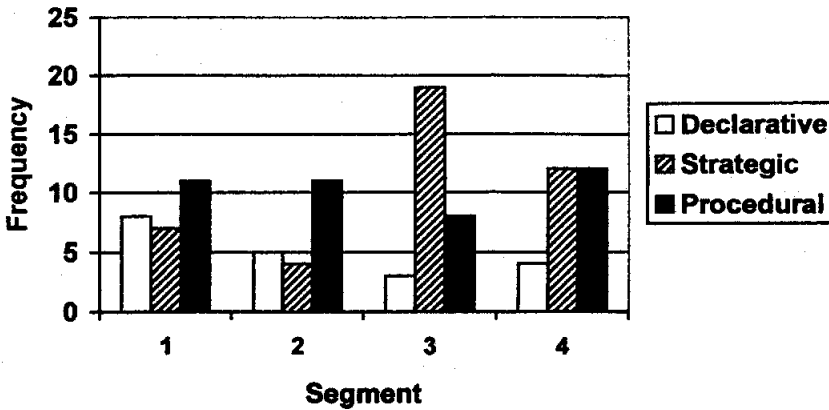


Figure 4. Frequencies of student responses coded into each mental model category for each segment.

Although task delivery was straightforward, the materials required to implement the method required a fair amount of preparation. This included study of the domain to identify knowledge elicitation goals, which in turn drove the scenario content. Thus before this technique can be used, it is likely that other knowledge elicitation techniques will have to be implemented. Preparation also included development of a script and subsequent development of videotape and support materials (e.g., questionnaire and briefs). From the standpoint of preparation of materials, then, this technique may require more resources than do other techniques.

**Data Format**

Although event-based knowledge elicitation may require more up-front preparation than

other techniques, the data analysis problem is greatly reduced. Hoffman (1987) defined the criterion *data format* as referring to whether data resulting from the task are in a format that can be readily entered in a database. Using event-based knowledge elicitation, because job contexts presented are known beforehand, expectations or hypotheses can be developed as well as approaches to coding and analysis. Event-based knowledge elicitation would seem to serve the purposes underlying data format, possibly more so than other approaches. The use of questionnaires in the present study made data readily codable and amenable for a database.

**Task Flexibility**

The flexibility of event-based knowledge elicitation may be one of its greatest strengths.

In the present study we were able to apply the task to members of the same community who varied in terms of their experience. We can envision its applications to many job applications, whether operating a nuclear power plant, flying an aircraft, controlling a large tactical team, or diagnosing X-rays. The critical requirement is whether the domain lends itself to scenario-based training and testing.

### Efficiency

Hoffman (1987) defined *efficiency* as the number of informative propositions produced per task minute; *task* includes preparation, knowledge elicitation, and analysis of the data. He estimated that unstructured interviews generate 0.13 propositions per task minute compared with the method of tough cases, which generates between one and two propositions per task minute.

Efficiency was not calculated for the present effort. However, it can be surmised that the task will be comparable to the method of tough cases, given that scenario events are designed to focus the responses on specific types of information and that data can be collected from more than one expert at a time.

### Task Artificiality

*Task artificiality* refers to the extent to which knowledge can be elicited on tasks that occur in the experience of the expert (Hoffman, 1987). As with approaches such as the method of tough cases, special, infrequently occurring job situations may be presented to experts in order to reveal aspects of problem-solving approaches that may not be apparent under nominal conditions.

### Validity

Finally, *validity*, as defined by Hoffman (1987), refers to whether the data resulting from the technique provides correct and important information about expert knowledge and reasoning. The results suggest that the methodology will result in differences among respondents based on their experience along the lines of what would be expected based on novice-expert differences, a finding that would seem important for any knowledge elicitation technique. It is also a technique, as implement-

ed for the present research, that lends itself to collecting responses from many respondents on similar, meaningful job situations. This can ensure completeness of the data and enable the assessment of expert agreement. The real key to obtaining valid data with the technique is the up-front analysis that results in targeted scenario events.

## CONCLUSION

In conclusion, using event-based knowledge elicitation, we obtained differences between respondents based on their experience level, providing some empirical evidence for the validity of the approach. The technique requires significant preparation time compared with other approaches, but because expectations can be developed, it may greatly simplify the data analysis. This is especially true if focused information is sought and the analyst plans to collect data from a number of experts.

Besides fulfilling a role in the knowledge elicitation process, other applications are possible. Because differences were obtained based on experience, a variation of the technique may serve a training purpose. We can also envision a testing application. For example, a point made by Zachary, Ryder, and Hicinbothom (in press) is that over time, there may be differences in the way experts perform a task as a result of factors such as the introduction of new tactics, techniques, and procedures and the introduction of equipment changes. Thus event-based knowledge elicitation could be used to establish a baseline of expert responses and then to periodically test operators to determine if there have been changes in the way the job tasks are being performed. Such testing may be necessary, for example, if decision aids or expert systems are being used. These and other applications could be explored in future efforts.

## ACKNOWLEDGMENTS

The views expressed here are those of the authors and do not necessarily represent the official positions of the agencies with which they are affiliated. We thank Nancy Cooke and Laura Milham for their assistance and advice regarding the coding taxonomies.

## REFERENCES

- Cannon-Bowers, J. A., Salas, E., & Converse, S. A. (1995). Shared mental models in expert team decision making. In N. J. Castellan, Jr. (Ed.), *Individual and group decision making: Current issues* (pp. 221-246). Mahwah, NJ: Erlbaum.
- Converse, S. A., & Kahler, S. E. (1992). *Knowledge acquisition and the measurement of shared mental models*. Unpublished manuscript, Naval Training Systems Center, Orlando, FL.
- Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies*, 41, 801-849.
- Duda, R. O., & Shortliffe, E. H. (1983). Expert systems research. *Science*, 220, 261-268.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32-64.
- Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, 47, 275-305.
- Fowlkes, J., Dwyer, D. J., Oser, R. L., & Salas, E. (1998). Event-based approach to training (EBAT). *International Journal of Aviation Psychology*, 8, 209-221.
- Fowlkes, J. E., Lane, N. E., Salas, E., Franz, T., & Oser, R. (1994). Improving the measurement of team performance: The TARGETs methodology. *Military Psychology*, 6, 47-61.
- Hoffman, R. R. (1987). The problem of extracting the knowledge of experts from the perspective of experimental psychology. *AI Magazine*, 8, 53-67.
- Hoffman, R. R., Shadbolt, N. R., Burton, A. M., & Klein, G. (1995). Eliciting knowledge from experts: A methodological analysis. *Organizational Behavior and Human Decision Processes*, 62, 129-158.
- Johnston, J. H., Smith-Jentsch, K. A., & Cannon-Bowers, J. A. (1997). Performance measurement tool for enhancing team decision-making training. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 311-327). Mahwah, NJ: Erlbaum.
- Rasmussen, J. (1986). *Information processing and human machine interaction: An approach to cognitive engineering*. New York: Elsevier.
- Redding, R. E., Cannon, J. R., & Seamster, T. L. (1992). Expertise in air traffic control (ATC): What is it, and how can we train for it? In *Proceedings of the Human Factors Society 36th Annual Meeting* (pp. 1362-1370). Santa Monica, CA: Human Factors and Ergonomics Society.
- Salas, E., Dickinson, T. L., Converse, S., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3-29). Norwood, NJ: Ablex.
- Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. *Human Factors*, 37, 123-136.
- Stout, R. J., Cannon-Bowers, J. A., & Salas, E. (1994). The role of shared mental models in developing shared situational awareness. In R. D. Gilson, D. J. Garland, & J. M. Koonce (Eds.), *Situational awareness in complex systems* (pp. 297-304). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.
- Zachary, W. W., Ryder, J. M., & Hicinbothom, J. H. (in press). Building cognitive task analyses and models of a decision making team in a complex real-time environment. In J. M. Schraagen, S. F. Chipman, & V. Shalin (Eds.), *Cognitive task analysis*. Mahwah, NJ: Erlbaum.
- Jennifer E. Fowlkes is a research psychologist in the Team Performance Laboratory at the University of Central Florida. She received her Ph.D. in experimental psychology from the University of Georgia in 1990.
- Eduardo Salas is a professor of psychology at the University of Central Florida and principal scientist for human factors research at the Institute for Simulation Training. He received his Ph.D. in 1984 from Old Dominion University, Norfolk, Virginia.
- David P. Baker is a senior research psychologist at the American Institutes for Research. He received his Ph.D. in industrial/organizational psychology from the University of South Florida in 1991.
- Janis A. Cannon-Bowers is a senior research psychologist at the Naval Air Warfare Center Training Systems Division. She received her Ph.D. in industrial/organizational psychology from the University of South Florida in 1988.
- Renée J. Stout is a performance development manager at Alignmark, Orlando, FL. She received her Ph.D. in human factors psychology from the University of Central Florida in 1994.