Ergonomics, 1998, vol. 41, no. 11, 1618–1641

Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands

Laura G. Militello and Robert J. B. Hutton

Klein Associates Inc., 582 E. Dayton-Yellow Springs Road, Fairborn, Ohio 45324, USA

Keywords: Cognitive task analysis; Knowledge elicitation; Knowledge representation; Expertise; Validity; Reliability.

Cognitive task analysis (CTA) is a set of methods for identifying cognitive skills, or mental demands, needed to perform a task proficiently. The product of the task analysis can be used to inform the design of interfaces and training systems. However, CTA is resource intensive and has previously been of limited use to design practitioners. A streamlined method of CTA, Applied Cognitive Task Analysis (ACTA), is presented in this paper. ACTA consists of three interview methods that help the practitioner to extract information about the cognitive demands and skills required for a task. ACTA also allows the practitioner to represent this information in a format that will translate more directly into applied products, such as improved training scenarios or interface recommendations. This paper will describe the three methods, an evaluation study conducted to assess the usability and usefulness of the methods, and some directions for future research for making cognitive task analysis accessible to practitioners. ACTA techniques were found to be easy to use, flexible, and to provide clear output. The information and training materials developed based on ACTA interviews were found to be accurate and important for training purposes.

1. Introduction

Task analytic techniques have played a critical role in the development of training and system design for the past 100 years starting with the pioneering work of Taylor (1911) and the Gilbreths (Gilbreth 1911, Gilbreth and Gilbreth 1919). (For more recent reviews of task analytic techniques see Drury *et al.* 1987, Kirwan and Ainsworth 1992). As task analytic techniques have become more sophisticated, focusing on cognitive activities as well as behaviours, they have become less accessible to practitioners. This paper introduces Applied Cognitive Task Analysis (ACTA), a set of streamlined cognitive task analysis tools that have been developed specifically for use by professionals who have not been trained in cognitive psychology, but who do develop applications that can benefit from the use of cognitive task analysis.

Cognitive task analysis methods focus on describing and representing the cognitive elements that underlie goal generation, decision making, judgements, etc. Although cognitive task analyses often begin with high-level descriptions of the task based on observations or initial interviews, the bulk of the data collection occurs via in-depth interviews with subject matter experts. These interviews focus on gaining information about the cognitive strategies used to accomplish the task including situation assessment strategies, identification and interpretation of critical cues,

0014-0139/98 $12.00 © 1998 Taylor & Francis Ltd

metacognitive strategies, important perceptual distinctions, etc. One strength of these methods is that they aid experts in articulating knowledge that is generally difficult to verbalize. Although researchers are often cautioned concerning reliance on verbal protocol data (Nisbett and Wilson 1977), it can be argued that verbal reports are inherently no more or less problematic than any other methodology (Howard 1994, Spector 1994).

The urgency of incorporating cognitive elements of performance into training and systems design stems from the changing nature of the workplace and the impact of technology on many tasks and functions. Tasks today place greatly increased demands on the cognitive skills of workers. Howell and Cooke (1989) have argued that with advances in technology and machine responsibility, we have increased, rather than lowered, cognitive demands on humans. More procedural or predictable tasks are handled by smart machines, while humans have become responsible for tasks that require inference, diagnoses, judgement, and decision making. In manufacturing technologies, for example, key skills now include perceptual skills for monitoring equipment, diagnostic skills for interpreting computerized information, and communication skills required for problem solving and co-ordination in distributed decision environments. These skills are typically grounded in hands-on experience with a task and are among the most difficult for proficient performers to access or articulate. Technology also has the potential to simplify low-level jobs; however, the concern shared by many is that it makes high-level jobs even more complex.

While a wide range of powerful methods of cognitive task analysis have been developed and applied over the last 10 years, few have become accessible to training practitioners and the engineering community designing systems. These have been described in several different sources (Klein 1993, Cooke 1994, Seamster *et al.* 1997). For example, in a multi-year effort Hall *et al.* (1994) used the Precursor, Action, Result, and Interpretation (PARI) method to develop an avionics troubleshooting tutor. Roth *et al.* (1991) used a cognitive environment simulation to investigate the cognitive activities involved in fault management with nuclear power plant operators. Seamster *et al.* (1993) report conducting an extensive cognitive task analysis to specify the instructional content and sequencing for a US Federal Aviation Administration *en route* air traffic control curriculum. Rasmussen (1986) has conducted a thorough cognitive analysis of nuclear power plant control room operation. Rouse (1984) has used similar methods to derive problem-solving strategies for troubleshooting tasks in the military.

Although these success stories are very persuasive in terms of the power of cognitive task analysis, all have required considerable time and resources. All have been part of research efforts conducted by scientists as opposed to the development of an application by practitioners. The objective of the authors has been to transition CTA techniques from the research community to the operational community. Based on existing CTA techniques, the authors have developed streamlined CTA methods intended for use by instructional designers and systems designers rather than knowledge engineers, cognitive psychologists, and human factors/ergonomics professionals. This paper describes Applied Cognitive Task Analysis (ACTA), streamlined CTA methods developed for training practitioners and systems designers to elicit and represent cognitive components of skilled task performance, and the means to transform those data into design recommendations.

2. Applied Cognitive Task Analysis

The ACTA techniques were developed as part of a two-year project funded by the Navy Personnel Research and Development Center. The goal of this project was to develop and evaluate techniques that would enable instructional designers and systems designers to elicit critical cognitive elements from Subject Matter Experts (SMEs). The techniques presented here are intended to be complementary; each is designed to get at different aspects of cognitive skill. The first technique, the task diagram interview, provides the interviewer with a broad overview of the task and highlights the difficult cognitive portions of the task to be probed further with in-depth interviews. The second technique, the knowledge audit, surveys the aspects of expertise required for a specific task or subtask. As each aspect of expertise is uncovered, it is probed for concrete examples in the context of the job, cues and strategies used, and why it presents a challenge to inexperienced people. The third technique, the simulation interview, allows the interviewer to probe the cognitive processes of the SMEs within the context of a specific scenario. The use of a simulation or scenario provides job context that is difficult to obtain via the other interview techniques, and therefore allows additional probing around issues such as situation assessment, how situation assessment impacts a course of action, and potential errors that a novice would be likely to make given the same situation. Finally, a cognitive demands table is offered as a means to consolidate and synthesize the data, so that it can be directly applied to a specific project. Each technique is described in detail in the following sections.

2.1. *Task diagram*

The task diagram elicits a broad overview of the task and identifies the difficult cognitive elements. Although this preliminary interview offers only a surface-level view of the cognitive elements of the task, it enables the interviewer to focus the more in-depth interviews (i.e. the knowledge audit and simulation interviews) so that time and resources can be spent unpacking the most difficult and relevant of those cognitive elements.

The subject matter expert is asked to decompose the task into steps or subtasks with a question such as, 'Think about what you do when you (*task of interest*). Can you break this task down into less than six, but more than three steps?' The goal is to get the expert to walk through the task in his/her mind, verbalizing major steps. The interviewer limits the SME to between three and six steps, to ensure that time is not wasted delving into minute detail during the surface-level interview. After the steps of the task have been articulated, the SME is asked to identify which of the steps require cognitive skill, with a question such as, 'Of the steps you have just identified which require difficult cognitive skills? By cognitive skills I mean judgements, assessments, problem solving–thinking skills'. The resulting diagram (figure 1) serves as a road map for future interviews, providing an overview of the major steps involved in the task and the sequence in which the steps are carried out, as well as which of the steps require the most cognitive skill.

The task diagram interview is intended to provide a surface-level look at the task, and does not attempt to unpack the mental model of each SME. The goal is to elicit a very broad overview of the task. Efforts to delineate a mental model can quickly degenerate into a search for everything in a person's head as Rouse and Morris (1986) have pointed out. In this interview, the authors recommend that the SME be limited to six steps and advise the interviewer not to get drugged down to a level of detail that is best captured by other interview techniques.

2.2. *Knowledge audit*

The knowledge audit identifies ways in which expertise is used in a domain and provides examples based on actual experience. The knowledge audit draws directly from the research literature on expert-novice differences (Dreyfus 1972, Chi *et al.* 1981, Shanteau 1985, Dreyfus and Dreyfus 1986, Hoffman 1992, Klein and Hoffman 1993) and critical decision method studies (Klein *et al.* 1989, Crandall and Getchell-Reiter 1993, Klinger and Gomes 1993, Militello and Lim 1995, Kaempf *et al.* 1996) of expert decision-making. The knowledge audit has been developed as a means for capturing the most important aspects of expertise while streamlining the intensive data collection and analysis methods that typify studies of expertise.

The knowledge audit is organized around knowledge categories that have been found to characterize expertise. These include; diagnosing and predicting, situation awareness, perceptual skills, developing and knowing when to apply tricks of the trade, improvising, metacognition, recognizing anomalies, and compensating for equipment limitations. Clearly, the authors could have included many more items, but the intent was to aim for the smallest number of high impact components.

The knowledge audit employs a set of probes designed to describe types of domain knowledge or skill and elicit appropriate examples (figure 2). The goal is not simply to find out whether each component is present in the task, but to find out the nature of these skills, specific events where they were required, strategies that have been used, and so forth. The list of probes is the starting point for conducting this interview. Then, the interviewer asks for specifics about the example in terms of critical cues and strategies of decision making. This is followed by a discussion of potential errors that a novice, less-experienced person might have made in this situation.

The examples elicited with the knowledge audit do not contain the extensive detail and sense of dynamics that more labour-intensive methods such as the critical decision method (Klein *et al.* 1989) incident accounts often do. However, they do provide enough detail to retain the appropriate context of the incident. It is not expected that all probes will be equally relevant in each domain. After a few interviews, interviewers can easily determine which probes have the highest payoff. Although the knowledge audit does not capture the depth of relationship of a conceptual graph structure (Gordon and Gill 1992) or other intensive methods, it does address a full range of aspects of expertise that are usually neglected by behavioural task analytic methods.

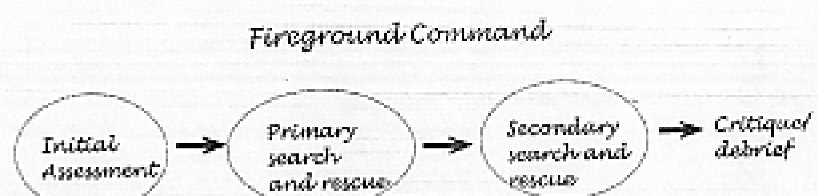

Figure 1. Task diagram example.

The output of the knowledge audit is shown in table 1, which contains an inventory of task-specific expertise. This table includes examples of situations in which experience has been called into play, cues and strategies used in dealing with these difficult situations, and an explanation of why such situations present a challenge to less-experienced operators.

2.3. *Simulation interview*

The simulation interview allows the interviewer to better understand the SME's cognitive processes within the context of an incident. In operational settings, the point of the job is typically to act upon the environment in some manner. Klein (1993) and

BASIC PROBES:

- **Past & Future.** Experts can figure out how a situation developed, and they can think into the future to see where the situation is going. Among other things, this can allow experts to head off problems before they develop.
  Is there a time when you walked into the middle of a situation and knew exactly how things got there and where they were headed?
- **Big Picture.** Novices may only see bits and pieces. Experts are able to quickly build an understanding of the whole situation—the Big Picture view. This allows the expert to think about how different elements fit together and affect each other.
  Can you give me an example of what is important about the Big Picture for this task? What are the major elements you have to know and keep track of?
- **Noticing.** Experts are able to detect cues and see meaningful patterns that less-experienced personnel may miss altogether.
  Have you had experiences where part of a situation just "popped" out at you; where you noticed things going on that others didn't catch? What is an example?
- **Job Smarts.** Experts learn how to combine procedures and work the task in the most efficient way possible. They don't cut corners, but they don't waste time and resources either.
  When you do this task, are there ways of working smart or accomplishing more with less—that you have found especially useful?
- **Opportunities/Improvising.** Experts are comfortable improvising—seeing what will work in this particular situation; they are able to shift directions to take advantage of opportunities.
  Can you think of an example when you have improvised in this task or noticed an opportunity to do something better?
- **Self Monitoring.** Experts are aware of their performance; they check how they are doing and make adjustments. Experts notice when their performance is not what it should be (this could be due to stress, fatigue, high workload, etc.) and are able to adjust so that the job gets done.
  Can you think of a time when you realized that you would need to change the way you were performing in order to get the job done?

OPTIONAL PROBES:

- **Anomalies.** Novices don't know what is typical, so they have a hard time identifying what is atypical. Experts can quickly spot unusual events and detect deviations. And, they are able to notice when something that ought to happen, doesn't.
  Can you describe an instance when you spotted a deviation from the norm, or knew something was amiss?
- **Equipment Difficulties.** Equipment can sometimes mislead. Novices usually believe whatever the equipment tells them; they don't know when to be skeptical.
  Have there been times when the equipment pointed in one direction, but your own judgment told you to do something else? Or when you had to rely on experience to avoid being led astray by the equipment?

Figure 2. Knowledge audit probes.

Howell and Cooke (1989) have asserted that identification and exploration of information surrounding high consequence, difficult decisions can provide a sound basis for generation of effective training and systems design. Simulation- and incident-based interviews have been used successfully in many domains (Flanagan 1954, Grover 1983, Clarke 1987, Diederich *et al.* 1987, Bell and Hardiman 1989, Cordingley 1989, Klein *et al.* 1989, Thordsen *et al.* 1992, Hall *et al.* 1994).

The simulation interview is based on presentation of a challenging scenario to the SME. The authors recommend that the interviewer retrieves a scenario that already exists for use in this interview. Often, simulations and scenarios exist for training purposes. It may be necessary to adapt or modify the scenario to conform to practical constraints such as time limitations. Developing a new simulation specifically for use in the interview is not a trivial task and is likely to require an upfront CTA in order to gather the foundational information needed to present a challenging situation. The simulation can be in the form of a paper-and-pencil exercise, perhaps using maps or other diagrams. In some settings it may be possible to use video or computer-supported simulations. Surprisingly, in the authors' experience, the fidelity of the simulation is not an important issue. The key is that the simulation presents a challenging scenario.

After exposure to the simulation, the SME is asked to identify major events, including judgements and decisions, with a question such as, 'As you experience this simulation, imagine you are the (*job you are investigating*) in the incident. Afterwards, I am going to ask you a series of questions about how you would think and act in this situation'. Each event is probed for situation assessment, actions, critical cues, and potential errors surrounding that event (figure 3).

Table 1. Example of a knowledge audit table.

Aspects of expertise	Cues and strategies	Why difficult?
Past and future e.g. Explosion in office strip; search the office areas rather than source of explosion	Material safety data sheets (MSDS) tell you that explosion in area of dangerous chemicals and information about chemicals Start where most likely to find victims and own safety considerations	Novice would be trained to start at source and work out May not look at MSDS, to find potential source of explosion, and account for where people are most likely to be
Big picture Big picture includes source of hazard, potential location of victims, ingress/egress routes, other hazards	Senses, communication with others, building owners, MSDS, building pre-plans	Novice gets tunnel vision, focuses on one thing, e.g. victims
Noticing Breathing sounds of victims	Both you and partner stop, hold your breath, and listen Listen for crying, talking to themselves, victims knocking things over	Noise from own breathing in apparatus, fire noises Don't know what kinds of sounds to listen for

Information elicited is recorded in the simulation interview table (table 2). Using the same simulation for interviews with multiple SMEs can provide insight into situations in which more than one action would be acceptable, and alternative assessments of the same situation are plausible. This technique can be used to highlight differing SME perspectives, which is important information for developing training and system design recommendations. The technique can also be used to contrast expert and novice perspectives by conducting interviews with people of differing levels of expertise using the same simulation.

For each major event, elicit the following information

As the (job you are investigating) in this scenario, what actions, if any, would you take at this point in time?

What do you think is going on here? What is your assessment of the situation at this point in time?

What pieces of information led you to this situation assessment and these actions?

What errors would an inexperienced person be likely to make in this situation?

Figure 3. Simulation interview probes.

Table 2. Example of a simulation interview table.

Events	Actions	Assessment	Critical cues	Potential errors
On-scene arrival	Account for people (names) Ask neighbours (but don't take their word for it, check it out yourself) Must knock on or knock down to make sure people aren't there	It's a cold night, need to find place for people who have been evacuated	Night time Cold -> 15° Dead space Add on floor Poor materials wood (punk board), metal girders (buckle and break under fire) Common attic in whole building	Not keeping track of people (could be looking for people who are not there)
Initial attack	Watch for signs of building collapse If signs of building collapse, evacuate and throw water on it from outside	Faulty construction, building may collapse	Signs of building collapse include: What walls are doing: cracking What floors are doing: groaning What metal girders are doing: clicking, popping Cable in old buildings hold walls together	Ventilating the attic, this draws the fire up and spreads it through the pipes and electrical system

2.4 *Cognitive demands table*

After conducting ACTA interviews with multiple SMEs, the authors recommend the use of a cognitive demands table (table 3) to sort through and analyse the data. Clearly, not every bit of information discussed in an interview will be relevant for the goals of a specific project. The cognitive demands table is intended to provide a format for the practitioner to use in focusing the analysis on project goals. The authors offer sample headings for the table based on analyses that they have conducted in the past (difficult cognitive element, why difficult, common errors, and cues and strategies used), but recommend that practitioners focus on the types of information that they will need to develop a new course or design a new system. The table also helps the practitioner see common themes in the data, as well as conflicting information given by multiple SMEs.

3. Evaluation study

As Hoffman *et al.* (1998) point out, the question of how to empirically verify a knowledge base, and the methodologies used to articulate and represent that knowledge base, has received little attention from the research community. Many

Table 3. Example of a cognitive demands table.

Difficult cognitive element	Why difficult?	Common errors	Cues and strategies used
Knowing where to search after an explosion	Novices may not be trained in dealing with explosions. Other training suggests you should start at the source and work outward Not everyone knows about the Material Safety Data Sheets. These contain critical information	Novice would be likely to start at the source of the explosion. Starting at the source is a rule of thumb for most other kinds of incidents	Start where you are most likely to find victims, keeping in mind safety considerations Refer to Material Safety Data Sheets to determine where dangerous chemicals are likely to be Consider the type of structure and where victims are likely to be Consider the likelihood of further explosions. Keep in mind the safety of your crew
Finding victims in a burning building	There are lots of distracting noises. If you are nervous or tired, your own breathing makes it hard to hear anything else	Novices sometimes don't recognize their own breathing sounds; they mistakenly think they hear a victim breathing	Both you and your partner stop, hold your breath, and listen Listen for crying, victims talking to themselves, victims knocking things over, etc.

cognitive task analysis methods are evaluated solely on the basis of subjective judgements of whether or not they seemed to work for a particular application or project. Exceptions include work by Crandall and her colleagues in assessing the validity and reliability of data-gathering skills using the critical decision method (Taynor *et al.* 1987, Crandall and Calderwood 1989, Crandall and Gamblian 1991, Crandall and Getchell-Reiter 1993), and method comparisons by Hoffman and his colleagues (Hoffman 1987, Hoffman *et al.* 1995). The evaluation study described in this section not only attempts to address issues of validity and reliability for a specific set of CTA techniques, but also addresses a number of issues that surround the assessment of validity and reliability within the context of real-world tasks.

3.1. *Methods*

An evaluation of the ACTA techniques was conducted to establish the validity and reliability of the data gathered using the methods, and to assess the usability of the techniques. In addition, a comparison of information gathered using ACTA techniques to data gathered using unstructured interview techniques was conducted.

Parallel studies were conducted in two domains for this evaluation. The authors' intention was to test the ACTA techniques with a sample of naive users—people who lacked knowledge or experience with cognitive task analysis or instructional design. A novice sample would allow a cleaner examination of the impact of the ACTA methods on the kind and quality of data produced. Therefore, students from graduate programmes in clinical, human factors, or cognitive psychology were recruited via postings on college bulletin boards and e-mail, to conduct interviews and generate instructional materials in either the firefighting domain or the Electronic Warfare (EW) domain. Volunteers were screened to make sure that they had no previous knowledge of the domain they would be investigating, no previous experience in conducting CTA, and no extensive experience or training in developing course materials. Each student was paid US $250 for participation in the project. Twelve students conducted interviews in the firefighting domain and 11 in the EW domain. The SMEs interviewed were experienced Fireground Commanders from the greater Dayton area in Dayton, Ohio, USA, who had at least 10 years of experience and were recommended by the Fire Chief of each local fire department; and experienced EW technicians from Fleet Training Center Pacific in San Diego, California, USA, who had at least 6 years of experience as EW technicians, including 4 years at sea and experience as an EW supervisor.

Within each domain, students were placed in one of two groups. An attempt was made to match the groups by age, gender, and educational level. After matching the students on these criteria, they were randomly assigned to groups. All students attended a 2-h workshop introducing the concepts underlying cognitive task analysis, describing the application of cognitive task analysis to the development of instructional materials, and providing a brief overview of the domain and specific task that they would be investigating. They also received instruction regarding the training materials they would be asked to develop following their interviews with SMEs. These materials included a cognitive demands table, learning objectives, and modifications to a training manual.

After the initial 2-h workshop, the matched groups of students were separated for the remainder of the study. One group, referred to as the Unstructured group, was provided with instructions to conduct interviews with SMEs in whatever format they believed would be most useful for gathering cognitive information. They were told to spend time preparing questions, but were not given any direction regarding how to structure the interviews or specific types of questions to ask. The other group, referred to as the ACTA group, was provided with a 6-h workshop on the ACTA techniques, including knowledge elicitation and knowledge representation.

Students in both the Unstructured and ACTA groups then participated in two interviews with SMEs. Each student led one interview with an SME and observed an interview conducted by another student with a different SME, and thus had access to data from two interviews. Each student in the ACTA group led one ACTA interview and observed an ACTA interview in the same domain; each student in the Unstructured group led one unstructured interview and observed one unstructured interview in the same domain. No SME was interviewed twice. Students working in the firefighting domain were asked to focus on the size-up task. Students working in the EW domain were asked to focus on signal threat analysis. All interviews were scheduled for a 3-h block of time.

Within a week of completing the two interviews, each student attended a 4-h session to analyse the data and develop training materials. The students were instructed not to collaborate or do any additional work with their interview notes prior to the scheduled session. During the 4-h session, they were required to structure and represent the information obtained in interviews. They were provided with materials and instructions and asked to:

(1) consolidate the data from the interview using the cognitive demands table format;
(2) develop at least ten cognitive learning objectives for a hypothetical course in that domain;
(3) revise or add to training manuals (these were provided), based on what they had learned in the interviews; and
(4) complete a questionnaire about participation in the study.

In addition, all students who had been exposed to the ACTA techniques were asked to fill out an ACTA usability questionnaire.

3.2. *Data transformation*

In order to generate quantitative measure of utility and validity, the information generated by the ACTA tools required extensive data codification and transformation. All materials generated by the sample of graduate students were assessed by SMEs who had not participated in the study thus far and/or cognitive psychologists. Wherever possible, data evaluation was carried out by multiple coders, so that inter-rater reliability could be assessed. In some cases, owing to lack of availability or resource constraints, only one SME was available to code the data. Measures were devised to address two aspects of validity: (1) whether the ACTA tools produced information that was predominantly cognitive in nature; and (2) whether the information produced was domain-specific and relevant. Data transformation procedures and associated measures are described in detail in the following sections.

3.2.1. *Validity indices—cognitive demands tables*: All items included in the cognitive demands tables were coded by two Klein Associates researchers, Laura Militello and Dr Rebecca Pliske, blind to treatment group (ACTA versus Unstructured), for whether they contained cognitive content. The criterion for inclusion in the cognitive demand category was that the item addressed a cognitive skill or a cognitive challenge that a firefighter/EW operator encounters on the job (e.g. 'deciding whether or not water supply on the scene will be adequate', 'devising a strategy to successfully remove people from a burning building'). Items in the non-cognitive demands category typically referred to declarative knowledge that the firefighter/EW operator should have (e.g. 'know the initial command sequence') or behaviours (e.g. 'return resources'). [Although it could be argued that knowing the proper command sequence has a cognitive component, it does not constitute a cognitive challenge or skill needed to serve as a proficient Fireground Commander. Rather, it represents a type of background knowledge that one must obtain before becoming a Fireground Commander.]

In order to establish whether students using the ACTA techniques could consistently elicit information across relevant cognitive categories (as opposed to task-based categories), the authors developed a coding scheme based on Rasmussen's model of decision making (Rasmussen 1986, Rasmussen *et al.* 1994). The categories included information collection, situation analysis, diagnosis, prediction, value judgement, choice, planning, and scheduling. Two raters, blind to the students' interview group (ACTA versus Unstructured), independently rated 30% of the data. The raters established acceptable inter-rater agreement (percentage agreement = 74%). [Although no standard for acceptable inter-rater agreement exists (Meister 1985, Pedhazur and Schmelkin 1991), agreement ratings exceeding 70% are generally accepted as adequate for this type of coding. Subsets of data were analysed until an acceptable level of agreement was reached. The remaining data were then analysed by one coder.] The rest of the data were then divided among the two raters to complete the rating.

Evaluation of the domain-specific content of the cognitive demands tables was based on the firefighting portion of the database. A task-based coding scheme specific to the firefighting domain was developed. Based on firefighting manuals made available to the authors by the National Fire Academy, Emmitsburg, MD, USA, it was established that there are three primary subtasks for which the Fireground Commander is responsible: size-up, strategy/tactics, and management. For the firefighter data, the coders independently assessed the content of each cognitive demands table item and assigned it to one of these three categories. The coders established reliability (percentage agreement = 81%) on 40% of the data, and one researcher then coded the remainder of the data.

The authors believed that it was also important to have the data evaluated by domain experts, in order to assess data quality and relevance. An assessment of the firefighter data was carried out by an instructor for the Incident Command course at the Ohio Fire Academy. He had more than 10 years of firefighting experience, had served as a Fireground Commander, and is currently involved in the development of course materials for the firefighter courses taught at the Ohio Fire Academy. The EW SME was a retired US Navy Electronic Warfare technician who had extensive experience as an operator, a supervisor, and as an instructor. The SMEs were asked to indicate what percentage of the information contained in each cognitive demands table would be likely to be known only by experienced personnel. In addition, the SMEs were asked to indicate the percentage of information contained in each cognitive demands table that would be relevant for experienced, highly-skilled personnel (Fireground Commander/EW supervisor) as opposed to a person with little experience on the job (firefighter/new EW operator). Given that one objective

of the ACTA techniques is to elicit experience-based knowledge (as opposed to classroom knowledge, which is easily captured using other traditional techniques and disseminated via textbooks), the authors wanted to distinguish information that only an experienced person would know from that which people newly released from training would know. Both of these questions were intended to distinguish between information reflective of experience-based knowledge versus classroom knowledge.

3.2.2. *Validity indices—instructional materials:* In addition to the firefighter SME described above, a second instructor from the Ohio Fire Academy was recruited to provide ratings of the instructional materials generated by the students in the firefighting domain. He also had more than 10 years experience as a firefighter, had served as a Fireground Commander, and is currently involved in the development of course materials at the Ohio Fire Academy. In addition, two EW instructors from the Electronic Warfare 'A' School in Pensacola, Florida, USA, were recruited to rate the instructional materials generated by graduate students working in the EW domain. Both Electronic Warfare SMEs held a rank of E6 or above, had served as an Electronic Warfare supervisor, and had experience as an instructor at the EW 'A' School.

Working independently, the SMEs in each domain were asked to evaluate each learning objective and training manual modification for accuracy, importance, and whether or not it was currently included in the typical firefighter training/EW instructional programme. Validity ratings were made on a 3-point scale where 1 = not important, 2 = important, and 3 = very important. Accuracy ratings were made on a two-point scale where 1 = not accurate and 2 = accurate. In the firefighting domain, acceptable inter-rater agreement was obtained for the accuracy and importance ratings, but not for the rating of whether or not the information described in the learning objective was currently covered in the typical firefighter training course. [Owing to unacceptable reliability ratings, no further analyses were conducted on data relating to whether the information was currently covered in a course. Discussion with the firefighter SMEs revealed that they had experience in teaching different courses and therefore had different perspectives on what was 'typically' included in firefighter instruction.] For the learning objectives, the percentage agreement for the accuracy judgements of the firefighter SMEs was 87.8%; the percentage agreement for the importance ratings of the firefighter SMEs was 71.4%. For the modifications to the student manual, percentage agreement for accuracy was 90.1% and for importance it was 76.1%. The accuracy and importance ratings made by the SME who had more extensive experience in developing training materials for Fireground Commanders were used in further analyses.

The SMEs in the EW domain were not able to reach acceptable inter-rater agreement. For the learning objectives, percentage agreement for importance was 58.5% and for accuracy it was 67.9%. For the modifications to the student manual, percentage agreement for importance was 34.2% and for accuracy it was 61.7%. Discussion with the SMEs revealed that depending upon the type of ship one serves on, the EW job may be very different. The two SMEs had served on different types of ships and were currently teaching very different courses (basic tactics versus introductory course on equipment). As a result, they had quite different perspectives on what was important for an EW operator to learn in school to prepare him/her for the job. For all further analyses, the authors used the ratings from the SME with the most recent and most extensive sea experience.

3.3 *Results*

The results section first presents the authors' findings as they relate to the usability, validity, and reliability of the ACTA techniques, as these were the primary questions to be answered by the evaluation study. Thus the data presented in the following sections are based only on students who completed the ACTA workshops and used these methods to conduct an interview with an SME. The final portion of the results section discusses the data as they relate to differences between the materials generated by students who conducted interviews using ACTA versus those students who conducted unstructured cognitive interviews. Although few group differences were found, a discussion of how large intra-group variability impacted on this study is presented.

3.3.1. *Usability*: In evaluating the ACTA tools, the authors were interested in understanding the subjective experiences of both the interviewers and the interviewees. User acceptance was key to the success of this project. To assess user acceptance, three questionnaires were administered: a usability questionnaire focusing specifically on the ACTA techniques, an interviewee questionnaire eliciting information from the SME's perspective, and an interviewer questionnaire addressing the experience of participating in all aspects of the study. The findings from the questionnaire data are presented next.

3.3.1.1. *Usability questionnaire*: A usability questionnaire was administered to all graduate students who used the ACTA techniques. Overall, ratings were very positive. All of the tools were rated as useful. A 3-way, mixed design ANOVA taking into account the domain (Firefighting, Electronic Warfare), the ACTA techniques (Task Diagram, Knowledge Audit, Simulation Interview), and the individual questions (Questions 2, 3, 4, and 5) on the usability questionnaire showed no difference in the usability of the three techniques, $F(2, 18) = 1.34$, $p = 0.29$, or in the usability of the techniques across domains, $F(1, 9)$, $p < 1$. Mean ratings on all dimensions were above '3' on a 5-point scale, where '5' is the most positive rating and '1' is the least positive rating (table 4). These data indicate that graduate students found:

(1) the methods to be easy to use;
(2) the interview guides and job aids to be flexible;
(3) the output of the interviews to be clear; and
(4) the knowledge representations to be useful.

3.3.1.2. *Interviewee questionnaire*: Each SME was asked to fill out a brief questionnaire at the end of the interview. If the ACTA tools are to be accepted in an operational community, the impressions of the people who are interviewed will have considerable influence. If the SMEs find the interview process aversive, or do not find that they are given an opportunity to communicate the critical elements of the job, acceptance of the ACTA tools will be greatly compromised within an organization.

The questionnaire data indicate that the interviewees found the interview experience to be pleasant and worthwhile. Table 5 presents the means for each question for those SMEs who participated in ACTA interviews. A 3-way, mixed design ANOVA taking into account domain (Firefighting, Electronic Warfare), interview type (ACTA, Unstructured), and question (5 questions from question-

Table 4. Usability questionnaire means of the graduate students who conducted interviews using the ACTA techniques.

Questions	Task diagram: Fire-fighting ($n=6$)	Task diagram: Electronic warfare ($n=5$)	Knowledge audit: Fire-fighting ($n=6$)	Knowledge audit: Electronic warfare ($n=5$)	Simulation interview: Fire-fighting ($n=6$)	Simulation interview: Electronic warfare ($n=5$)	ACTA overall: Fire-fighting ($n=6$)	ACTA overall: Electronic warfare ($n=5$)
Circle the amount of time you spent becoming familiar with the tool before using it†	1.00 (0.00)	1.40 (0.55)	1.17 (0.41)	1.40 (0.55)	1.00 (0.00)	1.20 (0.45)	NA	NA
Rate the degree to which you found this technique easy to use	4.67 (0.52)	4.60 (0.89)	3.67 (1.03)	3.80 (1.30)	3.83 (1.60)	4.40 (0.89)	4.17 (0.75)	4.20 (0.45)
Rate the degree to which you found the interview guide to be flexible	4.17 (0.75)	3.40 (0.55)	3.67 (1.03)	3.40 (0.55)	3.83 (1.17)	3.60 (1.14)	3.67 (1.21)	3.80 (0.45)
Rate the degree to which you found the output to be clear	4.67 (0.52)	4.00 (1.00)	4.00 (1.10)	3.60 (0.55)	3.83 (1.83)	4.60 (0.55)	4.00 (1.10)	4.00 (0.00)
Rate the degree to which you found the knowledge representation to be useful	4.50 (0.55)	4.00 (0.71)	4.17 (1.17)	3.80 (0.84)	3.33 (1.21)	4.00 (0.71)	4.33 (1.21)	4.40 (0.55)

*All questions used a 5-point rating scale where 5 is the most positive rating and 1 is the least positive rating, unless otherwise stated. Standard deviations are included in parentheses.
†A 3-point scale where 1 = 1 h or less; 2 = 1 to 2 h; and 3 = more than 2 h was used for the questions concerning time required to prepare for each tool.

Table 5. Interviewee questionnaire means for the SMEs interviewed using the ACTA techniques.

Questions	Firefighting ($n=6$)	Electronic warfare ($n=5$)	Mean totals ($n=11$)
Overall, I found the interview to be a pleasant experience	4.33 (0.52)	3.60 (1.14)	3.97 (0.89)
The format of the interview allowed me to describe my expertise	4.50 (0.55)	3.60 (1.14)	4.05 (0.94)
I thought the interview lasted too long† (1 = strongly agree, 5 = strongly disagree)	4.00 (0.63)	4.00 (0.71)	4.00 (0.63)
Participating in the interview gave me new insights into the cognitive aspects of my job	3.67 (1.03)	3.20 (1.48)	3.44 (1.21)
I think the cognitive aspects of my job that were discussed during the interview are important things for a novice to learn	4.33 (0.52)	4.00 (1.22)	4.17 (0.87)

*All questions used a 5-point rating scale. Unless otherwise stated, 1 = strongly disagree, 5 = strongly agree. Standard deviations are included in parentheses.
†The scale for question 3 was switched so that a high score indicates a more positive response to be consistent with other questions on the questionnaire.

naire) showed no domain differences in the interviewees' responses, $F(1, 20) = 0.82$, $p < 1$.

3.3.1.3. *Interviewer questionnaire:* All graduate students filled out a questionnaire at the completion of their participation in the evaluation study. This questionnaire consisted of 10 questions intended to capture the interviewer's subjective experience, addressing issues such as confidence level, perceived difficulty or ease of the cognitive task analysis process, etc. The means for each question from the ACTA group are presented in table 6.

These data indicate that graduate students in both domains found the interviews to be informative and to provide cognitive information about the job domain. Based on the information learned via ACTA interviews, the graduate students found the development of a cognitive demands table and the generation of learning objectives to be easy. Participants indicated that they were able to make important revisions to the course materials provided. Surprisingly, participants responded negatively to the statement, 'I want to conduct more interviews because I still want more information'. The authors' only explanation for this is that because the participants

Table 6. Interviewer questionnaire means for the graduate students who conducted interviews using the ACTA techniques.

Questions*	Firefighting ($n=6$)	Electronic warfare ($n=5$)	Means total ($n=11$)
I felt confident in my ability to lead an interview	4.33 (0.82)	4.00 (0.00)	4.17 (0.60)
I learned more information from the interview I observed, than from the one I led	2.00 (0.89)	3.00 (1.23)	2.50 (1.13)
I felt I had sufficient information to revise the course materials	4.00 (0.89)	3.20 (1.30)	3.60 (1.12)
The interviews provided me with important information about the cognitive skills involved in this job domain	4.33 (0.52)	3.60 (1.14)	3.97 (0.89)
I wanted to conduct more interviews because I still wanted more information	2.33 (0.82)	3.20 (1.30)	2.77 (1.10)
The cognitive demands table was easy to fill out	4.00 (1.26)	3.40 (0.89)	3.70 (1.10)
It was easy to develop course objectives based on information specified in the cognitive demands table	4.33 (0.52)	3.20 (1.64)	3.77 (1.25)
I was able to use the information to make important changes in the course material	3.50 (0.84)	3.40 (1.14)	3.45 (0.93)
Overall, I found the interviews to be informative	4.83 (0.41)	4.60 (0.55)	4.72 (0.47)
Given the information, I found the revision of course materials straightforward	3.67 (0.82)	3.00 (1.23)	3.34 (1.03)

*All questions used a 5-point rating scale. Unless otherwise stated, 1 = strongly disagree, 5 = strongly agree. Standard deviations are included in parentheses.

were novices to cognitive task analysis, they did not anticipate the breadth and depth of knowledge that can be gained via these techniques.

3.3.2. *Validity:* Table 7 presents data that addresses three central questions regarding validity:

(1) Does the information gathered address cognitive issues?
(2) Does the information gathered deal with experience-based knowledge as opposed to classroom-based knowledge?
(3) Do the instructional materials generated contain accurate information that is important for novices to learn?

The extent to which the information elicited using ACTA was cognitive in nature was assessed by examining every item contained in the cognitive demands tables for its cognitive content. The cognitive content codings indicate that fully 93% of the items generated address cognitive issues. More specifically, in the firefighter study, 92% of the items were rated as cognitive and in the EW study 94% of the cognitive demand items generated by the students using ACTA were rated as cognitive.

To address the issue of whether the ACTA tool provided a means to elicit experience-based knowledge, SMEs were asked to make a global assessment of each cognitive demands table and to assign a percentage to each that reflected the proportion of information it contained that only highly-experienced personnel would be likely to know. The inference here is that such information is reflective of experience-

Table 7. Quality of outputs for graduate students who conducted interviews using the ACTA techniques.

Validity indicator	Firefighter ($n=6$)	Electronic warfare ($n=5$)
Percentage of total cognitive demands table items coded as cognitive	92%	94%
Proportion of cognitive demands table information experienced personnel likely to know, averaged across ACTA users	0.95 (0.05)	0.90 (0.09)
Proportion of cognitive demands table information relevant to a Fireground Commander/Electronic Warfare Supervisor, averaged across ACTA users	0.73 (0.10)	0.87 (0.10)
Proportion of student manual modifications rated as important or somewhat important, averaged across ACTA users	0.70 (0.47)	0.93 (0.26)
Proportion of learning objectives rated as important or somewhat important, averaged across ACTA users	0.95 (0.43)	0.83 (0.38)
Proportion of student manual modifications rated as accurate, averaged across ACTA users	0.89 (0.31)	0.65 (0.48)
Proportion of learning objectives rated as accurate, averaged across ACTA users	0.92 (0.24)	0.54 (0.50)

Standard deviations are included in parentheses.

based knowledge, as opposed to knowledge that is easily obtained in a classroom setting. The information that newly-trained personnel possess is more likely to have been acquired in a classroom setting rather than through lived experiences.

The findings offer strong support that the ACTA tools allowed students to elicit important and relevant domain information. In the firefighter study, the percentage of content of the cognitive demands tables that was judged to be information that only highly-experienced personnel would know averaged 95%. In the EW domain, the same assessment yielded an average of 90% across the ACTA group. In response to questions regarding information relevance, a substantial percentage of information in the cognitive demands tables ($M = 73\%$) was rated as relevant for a Fireground Commander. Similar results were found in the EW domain where a mean of 87% of the information in the cognitive demands tables was rated as relevant to an EW supervisor.

The third validity question focused on the importance and accuracy of the instructional materials generated by ACTA users. The measures included ratings of importance and accuracy by domain experts. The 3-point importance ratings were collapsed into a dichotomy, with 'important' or 'somewhat important' ratings combined into a single importance indicator. Accuracy had been assessed as a dichotomy (accurate versus not). Findings indicate that in both domains, content of instructional materials generated by ACTA was viewed as important domain information for novices to learn. In the firefighting domain, a mean of 70% of the instructional material modifications generated, and 95% of the learning objectives generated by each student were rated as important. In the EW domain, these means were 93% and 83%, respectively (table 7).

Accuracy evaluations were also high, particularly for the firefighting data. In the firefighting domain a mean 89% of the modifications to the student manuals were rated as accurate and 92% of the learning objectives were rated as accurate. In the EW domains, these means were 65% (modifications to the student manual) and 54% (learning objectives). The authors suspect that the lower accuracy ratings in the EW domain were due to the more technical nature of the domain. The environment in which an EW operator/supervisor works was so foreign to the graduate students that understanding and using the terminology and acronyms that EW operators/supervisors use to describe the equipment and environment presented additional difficulty. There were a number of cases in which the electronic warfare SMEs rating the data indicated that they knew what the student must have meant, but that the wording used was incorrect.

3.3.3. *Reliability*: There is no well-established metric or method for assessing the reliability of cognitive task analysis tools, and yet the issue is an important one. Briefly, the question is whether individuals using a particular technique are able to generate comparable information, so that the tools can be considered a source of consistent information, given the same (or similar) domain expert assessed at different points in time and/or by different knowledge engineers. This is a much simpler matter when one is dealing with highly-structured interview formats or scale items than when faced with textual knowledge representations. The authors sought to address the issue in several ways. One approach was to examine whether ACTA users consistently elicited the same types of cognitive information. Therefore, the authors examined the content of the cognitive demands generated by the students, to see whether they had generated similar information.

One set of analyses examined whether ACTA users had generated similar types or categories of cognitive information. This analysis utilized a coding scheme based on Rasmussen *et al.*'s (1994) model of decision making. Each item in each cognitive demands table was coded. In the firefighting domain, every cognitive demands table (100%) generated by the ACTA group contained information that had to do with situation analysis and planning. All but one of the cognitive demands tables (80%) contained data on information collection. Given that students were instructed to focus on the subtask of 'size-up', which consists of gathering relevant information in order to accurately assess the situation and develop a plan of action, it was concluded that students in this study were able to consistently elicit relevant cognitive information using the ACTA techniques.

The same analysis was carried out for the EW study. All ACTA users generated cognitive demands that included information about situation analysis and all but one collected data in the information collection category. The signal threat analysis task consists primarily of gathering the necessary information to maintain an accurate, current assessment of the situation at all times. Again the data indicate that the students consistently elicited relevant cognitive information using the ACTA tools. Data across the two domains suggest that students were able to consistently elicit comparable cognitive information using the ACTA techniques.

A second coding scheme, specific to the firefighting data, also indicated that students consistently elicited similar information. The Fireground Commander task was divided into three subtasks: size-up; strategy and tactics; and management. All the ACTA users obtained information in each of these categories. The bulk of the information gathered focused on the size-up task (62%), which is where the students were asked to focus their interviews. Thus it is concluded that, using ACTA, people were consistently able to get important cognitive information for the entire Fireground Commander task, with an emphasis on the size-up task.

The authors also attempted to assess the degree of overlap of specific items across the cognitive demands tables generated by ACTA users. This proved to be extremely difficult, because users had not been constrained in level of detail, phrasing, or specificity. One student might list as a cognitive demand 'look of the smoke' while another noted 'color and movement of smoke'. The levels of inference required by raters to judge the degree to which any two cognitive demands table items matched were similar or were different became unacceptable, and the analysis was abandoned.

However, an informal examination of the cognitive demands tables suggests that the graduate students did not, in most cases, generate identical cognitive demands. This is not surprising given the design of this study. In order to reduce intra-subject variability, the authors excluded from the study graduate students who had any experience in the domain for which they would be conducting the cognitive task analysis; this meant that all of the students were working at a disadvantage in conducting the CTA. When the ACTA tools are described to professional audiences, it is recommended that time is spent becoming familiar with a domain before interviews are conducted. In this case, the students were given a brief overview of the domain and the task that they would be studying. The limited time we had with the graduate students did not allow for the recommended level of familiarization with the domain. A second reason why the graduate students did not generate the same cognitive demands is that each student was exposed to only two interviews with SMEs. If SME availability had allowed each student access to three to five experts as is generally recommended, the students would have been more likely to have heard similar things in the interviews.

3.3.4. *Group differences*: One of the drawbacks of the evaluation study design was the limited sample size. The intensive workshop preparation necessary to train subjects in ACTA methods, the extensive coding and data transformation effort necessary to provide empirical evaluation data, and the limited number of available SMEs, made large samples simply beyond the time or resources available. Obviously, with the small group sizes available, the effects associated with membership in the ACTA group were going to have to be very strong to be discernable as statistically significant.

Nonetheless, the authors were surprised to find so few differences between the ACTA group and the Unstructured interview group in the data. In addition to the small sample size, large intra-group differences were found that appear also to account for the lack of statistically significant results. Although an attempt was made to match the groups for age, gender, and education level, considerable individual differences were found in the students' comfort level and ease in conducting interviews (as observed by the investigators). This resulted in large standard deviations on nearly all the comparative measures, making those findings that were statistically significant difficult to interpret. For example, in rating the evaluation study experience, graduate students in the ACTA group for both the firefighter and the EW study agreed more strongly with the statement, 'I felt confident in my ability to lead an interview' ($M = 4.18$, $SD = 0.60$) than the graduate students who conducted unstructured interviews ($M = 3.25$, $SD = 0.87$), $U = 28.5$, $p = 0.02$. [A Mann-Whitney U-test, which is free of variance assumptions, was used instead of a t-test because there was no variance in the responses from the ACTA students in the electronic warfare domain.] In the firefighter study, the ACTA group agreed more strongly with the statement, 'I felt I had sufficient information to revise the course materials' ($M = 4.00$, $SD = 0.89$) than the Unstructured group ($M = 2.67$, $SD = 1.03$), $t(10) = 2.39$, $p = 0.06$. These statistical analyses indicate that students trained to use ACTA felt more confident in conducting the interviews and were more confident that they had gathered sufficient information to revise the course materials than the Unstructured group. However, in looking at the large standard deviations, it becomes clear that some students in each group were confident, whereas others were not.

In other cases, the means indicate very little difference between the groups, but the standard deviations indicate considerable variance within the groups. For example, the means for the two groups are nearly identical in response to the question, 'Given the information, I found the revision of course materials to be straightforward'. However, the large standard deviations indicate that some people in each group found the revision of the course materials straightforward, but others did not (table 8). Given the small sample sizes used in this study, it is clear that these group difference comparisons are not very robust.

One potentially confounding factor in the design of the study was that during the introductory workshop, the Unstructured group was exposed to a lecture on cognitive elements, cognitive task analysis, and how to fill out a cognitive demands table before conducting interviews. Examination of the raw data suggests that some of the students in the Unstructured group may have used the cognitive demands table to structure their interviews, thus reducing the gap between the ACTA group

Table 8. Mean responses to Question 10 on the interviewer questionnaire.

	ACTA group		Unstructured group	
	Firefighting domain	Electronic warfare domain	Firefighting domain	Electronic warfare domain
Given the information, I found the revision of course materials straightforward	3.67 (0.82)	3.83 (1.47)	3.00 (1.23)	3.33 (1.03)

Responses refer to a 5-point rating scale, where, 1 = strongly disagree, 5 = strongly agree. Standard deviations are in parentheses.

and the Unstructured group. The implication for ACTA is that the cognitive demands table is a valuable tool for framing the kinds of information that the interviewer intends to elicit from the SME.

Although the authors considered using a control group that would receive no introductory workshop on cognition, this was found to be impractical given that they wanted to compare both the amount of cognitive information elicited in the interviews and the quality of the training materials produced. In order to make these comparisons, it was necessary to provide training in how to create a cognitive demands table and how to produce instructional materials to all of the participants in the study.

The high quality ratings (i.e. SME ratings of importance and accuracy) received by both interview groups indicate that an exposure to the concepts underlying cognitive task analysis and a description of how cognitive task analytic data can be applied to instructional materials, may play a large role in learning to conduct cognitive task analyses. Working only with this foundational material, in the absence of exposure to actual methodologies, some students in the Unstructured group were able to gather accurate, relevant cognitive information and develop useful instructional materials.

4. Discussion

The findings presented here indicate that, after a 6-h workshop introducing the ACTA techniques, graduate students were able to conduct interviews with SMEs and elicit important, accurate cognitive information that was easily translated into instructional materials. Furthermore, subjective usability data indicate that the graduate students found the techniques to be easy to use, flexible, and to provide clear output. The authors' belief is that professional instructional designers and systems designers will do even better than the graduate students, given that they will have more concrete goals for use of the cognitive information and more experience in generating applications.

Although an attempt has been made here to establish the reliability and validity of the ACTA methods, the authors are aware that no well-established metrics exist. The need to test CTA methods in real-world settings with real-world tasks greatly reduces the level of control that one has over the many sources of variability. Factors that are difficult to control include the fact that some people seem to be more predisposed to be good interviewers than others. In addition, some SMEs are more articulate and easier to focus than others. Given the variability among humans in both the interviewer and the SME roles, it will be important to answer such questions as: Does an SME report the same examples and the same details when asked the same question later in time? Do the CTA techniques elicit the same types of information when used by different interviewers? Do independent practitioners generate the same materials based on CTA interviews? Further work is need to establish meaningful metrics to assess the reliability and validity of CTA tools.

An important point to make here is that although the ACTA methods have been shown to elicit important, accurate cognitive information, the authors have yet to try to assess what is lost using these streamlined techniques. It is believed that a trade-off exists: the more streamlined and proceduralized CTA techniques become, the less powerful they are. It appears that the ACTA techniques gather less comprehensive information than more systematic techniques such as the PARI method (Hall *et al.* 1994) and Gordon and Gill's (1992) conceptual graph analysis, and that the information gathered is more superficial than that gathered using the critical decision method (Klein *et al.* 1989) or Rasmussen's (1986) cognitive analysis. In spite of the limitations of streamlined CTA procedures, the ACTA techniques provided graduate students with sufficient tools to identify key cognitive elements and develop useful training materials. Until better metrics exist, however, it will be difficult to objectively assess what is lost and what is gained via different techniques.

It is also important to point out the impact of a failure to investigate and incorporate cognitive issues in complex tasks requiring high degrees of cognitive skill. Despite the promise of automated and intelligent systems, the human decision-maker will always play a role in systems where uncertainty and ambiguity exist. The consequences of not training operators to acquire the cognitive skills required, or not designing systems to support human problem-solving and decision-making can be dire, as illustrated by disasters such as Three Mile Island, the USS Vincennes, and Kings Cross Station, to name but a few (Reason 1990).

The ACTA methodology was originally taught in a workshop format, which allowed for the workshop presenters to tailor the methods to the audience and add personal anecdotes to the instruction. This also meant that no two ACTA workshops were the same. This made research into the reliability of the methods even more difficult to evaluate. However, more recently the ACTA workshop has been produced on a compact disk-based, multimedia training tool (Militello *et al.* 1997). This tool provides the learner with the reasons for undertaking a cognitive task analysis, an introduction to cognition and expertise, and a tutorial on the three ACTA methods. This tool provides an opportunity to conduct a more controlled comparison of the effectiveness and reliability of the ACTA methodology. Future research issues include looking at the reliability and validity of the knowledge elicitation and representations, as well as of the end-product training and systems interventions.

Future research using ACTA should also explore the following areas: incorporating methods for analysing tasks that rely implicitly on multi-operator interactions and teamwork; improving the incorporation of ACTA into more traditional and systematic task analytic techniques; improving the knowledge representations so that they are more easily translated into training and system design interventions; and, improving the training methods, and system and interface

design interventions themselves, so that cognitive strengths and cognitive vulnerabilities can be better supported.

Acknowledgements

This research was carried out under contract number N66001-94-C-7034 from the Navy Personnel Research and Development Center, San Diego, CA, USA.

References

Bell, J. and Hardiman, R. J. 1989, The third role—the naturalistic knowledge engineer, in D. Diaper (ed.), *Knowledge Elicitation: Principles, Techniques, and Applications* (New York: Wiley), 49–85.

Chi, M. T. H., Feltovich, P. J. and Glaser, R. 1981, Categorization and representation of physics problems by experts and novices, *Cognitive Science*, **5**, 121–152.

Clarke, B. 1987, Knowledge acquisition for real-time knowledge-based systems, *Proceedings of the First European Workshop on Knowledge Acquisition For Knowledge Based Systems* (United Kingdom: Reading University), September.

Cooke, N. J. 1994, Varieties of knowledge elicitation techniques, *International Journal of Human-Computer Studies*, **41**, 801–849.

Cordingley, E. S. 1989, Knowledge elicitation techniques for knowledge-based systems, in D. Diaper (ed.), *Knowledge Elicitation: Principles, Techniques, and Applications* (New York: Wiley), 89–175.

Crandall, B. and Calderwood, R. 1989, Clinical assessment skills of experienced neonatal intensive care nurses. Final report, Klein Associates Inc., OH. Prepared under contract 1 R43 NR0191101 for The National Center for Nursing, NIH.

Crandall, B. and Gamblian, V. 1991, Guide to early sepsis assessment in the NICU. Instruction manual prepared for the Ohio Department of Development under the Ohio SBIR Bridge Grant program by Klein Associates Inc., OH.

Crandall, B. and Getchell-Reiter, K. 1993, Critical decision method: A technique for eliciting concrete assessment indicators from the 'intuition' of NICU nurses, *Advances in Nursing Sciences*, **16**, 42–51.

Diederich, J., Ruhmann, I. and May, M. 1987, KRITON: A knowledge acquisition tool for expert systems, *International Journal of Man-Machine Studies*, **25**, 29–40.

Dreyfus, H. L. 1972, *What Computers Can't Do: A Critique of Artificial Reason* (New York: Harper & Row).

Dreyfus, H. L. and Dreyfus, S. E. 1986, *Mind over Machine: The Power of Human Intuitive Expertise in the Era of the Computer* (New York: The Free Press).

Drury, C. G., Paramore, B., Van Cott, H. P., Grey, S. M. and Corlett, E. N. 1987, Task analysis, in G. Salvendy (ed.), *Handbook of Human Factors* (New York: Wiley), 370–401.

Flanagan, J. C. 1954, The critical incident technique, *Psychological Bulletin*, **51**, 327–358.

Gilbreth, F. B. 1911, *Motion Study* (New York: Van Nostrand).

Gilbreth, F. B. and Gilbreth, L. M. 1919, *Fatigue Study* (New York: Macmillan).

Gordon, H. W. and Gill, R. T. 1992, Knowledge acquisition with question probes and conceptual graph structures, in T. Laurer, E. Peacock and A. Graesser (eds), *Questions and Information Systems* (Hillsdale, NJ: Lawrence Erlbaum), 29–46.

Grover, M. D. 1983, A pragmatic knowledge acquisition methodology, *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, Karlsruhe, Germany, August, 436–438.

Hall, E. M., Gott, S. P. and Pokorny, R. A. 1994, A procedural guide to cognitive task analysis: the PARI method, unpublished manuscript, Brooks AFB, TX.

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. 1992, *The Psychology of Expertise: Cognitive Research and Empirical AI* (New York: Springer-Verlag).

Hoffman, R. R., Crandall, B. E. and Shadbolt, N. R. 1998, A case study in cognitive task analysis methodology: The critical decision method for the elicitation of expert knowledge. *Human Factors* (In press).

Hoffman, R. R., Shadbolt, N. R., Burton, A. M. and Klein, G. 1995, Eliciting knowledge from experts: A methodological analysis, *Organizational Behavior and Human Decision Processes*, 62, 129–158.

Howard, G. S. 1994, Why do people say nasty things about self-reports? *Journal of Organizational Behavior*, 15, 399–404.

Howell, W. C. and Cooke, N. J. 1989, Training the human information processor: A look at cognitive models, in I.L. Goldstein (ed.), *Training and Development in Work Organizations: Frontiers of Industrial and Organizational Psychology* (San Francisco: Jossey-Bass), 121–182.

Kaempf, G. L., Klein, G., Thordsen, M. L. and Wolf, S. 1996, Decision making in complex command-and-control environments, *Human Factors Special Issue*, 38, 220–231.

Kirwan, B. and Ainsworth, L. K. 1992, *A Guide to Task Analysis* (London: Taylor & Francis).

Klein, G. 1993, Naturalistic decision making—implications for design. SOAR 93-01 (Contract no. DLA900-88-0393) CSERIAC, OH. State-of-the-Art Report 93-01; Dayton, OH: Crew System Ergonomics Information Analysis Center.

Klein, G. A. and Hoffman, R. 1993, Seeing the invisible: perceptual/cognitive aspects of expertise, in M. Rabinowitz (ed.), *Cognitive Science Foundations of Instruction* (Hillsdale, NJ: Lawrence Erlbaum), 203–226.

Klein, G. A., Calderwood, R. and MacGregor, D. 1989, Critical decision method for eliciting knowledge, *IEEE Transactions on Systems, Man, and Cybernetics*, 19, 462–472.

Klinger, D. and Gomes, M. 1993, A cognitive systems engineering application for interface design, *Proceedings of the 37th Annual Meeting of the Human Factors Society* (Santa Monica, CA: Human Factors Society).

Meister, D. 1985, The two worlds of human factors, in R. E. Eberts and C. G. Eberts (eds), *Trends in Ergonomics/Human Factors II*, (New York: North-Holland), 3–11.

Militello, L. G. and Lim, L. 1995, Early assessment of NEC in premature infants, *Journal of Perinatal and Neonatal Nursing*, 9, 1–11.

Militello, L. G., Hutton, R. J. B. and Miller, T. 1997, [Computer software] *Applied Cognitive Task Analysis* (Fairborn, OH: Klein Associates Inc).

Nisbett, R. E. and Wilson, T. D. 1977, Telling more than we can know: Verbal reports on mental processes, *Psychological Review*, 84, 231–259.

Pedhazur, E. J. and Schmelkin, L. P. 1991, *Measurement, Design, and Analysis: An Integrated Approach* (Hillsdale, NJ: Lawrence Erlbaum).

Rasmussen, J. 1986, *Information Processing and Human-machine Interaction: An Approach to Cognitive Engineering* (New York: North-Holland).

Rasmussen, J., Pettersen, A. M. and Goodstein, L. P. 1994, *Cognitive Systems Engineering* (New York: Wiley).

Reason, J. 1990, *Human Error* (New York: Cambridge University Press).

Roth, E. M., Woods, D. D. and Pople, H. E., Jr. 1992, Cognitive simulation as a tool for cognitive task analysis, *Ergonomics*, 35, 1163–1198.

Rouse, W. 1984, Models of natural intelligence for fault diagnosis tasks: Implications for training and aiding of maintenance personnel, in Artificial Intelligence in Maintenance: Proceedings of the Joint Services Workshop, report no. AFHRL-TR-84-25, Air Force Human Resources Laboratory, Colorado, 193–212.

Rouse, W. B. and Morris, N. M. 1986, On looking into the black box: Prospects and limits on the search for mental models, *Psychological Bulletin*, 100, 349–363.

Seamster, T. L., Redding, R. E., Cannon, J. R., Ryder, J. M. and Purcell, J. A. 1993, Cognitive task analysis of expertise in air traffic control, *The International Journal of Aviation Psychology*, 3, 257–283.

Seamster, T. L., Redding, R. E. and Kaempf, G. L. 1997, *Applied Cognitive Task Analysis in Aviation* (Aldershot: Avebury Aviation).

Shanteau, J. 1985, Psychological characteristics of expert decision makers, *Applied Experimental Psychology Series*, 85 (Kansas State University, Kansas).

Spector, P. E. 1994, Using self-report questionnaires in OB research: A comment on the use of a controversial method, *Journal of Organizational Behavior*, 15, 385–392.

Taylor, F. W. 1911, *The Principles of Scientific Management* (New York: Harper).

Taynor, J., Crandall, B. and Wiggins, S. 1987, The reliability of the critical decision method, KATR-863(B)-87-07F. Prepared under contract MDA903-86-C-0170 for the US Army Research Institute Field Unit, Leavenworth, KS, by Klein Associates Inc., OH.

Thordsen, M. L., Militello, L. G. and Klein, G. A. 1992, Cognitive task analysis of critical team decision making during multiship engagements Prepared under subcontract RI-82264X for Williams Air Force Base, AZ, by Klein Associates Inc., OH.