4. SOME GUIDELINES FOR CONDUCTING A COGNITIVE TASK ANALYSIS

Gary Klein and Laura Militello

ABSTRACT

Cognitive Task Analysis (CTA) attempts to explain the mental processes involved in performing a task. These processes include the knowledge, skills and strategies that are needed to accomplish the task functions. The criteria for success in a CTA study are: making a useful discovery about the cognitive skills being studied; being able to communicate the discovery to the users (i.e. those who will need to use the CTA for design); and having a meaningful impact on the eventual design.

Currently, a wide variety of CTA methods are being used. As we learn how to define the cognitive demands presented by a task/situation, we hope we will be able to map CTA methods onto these demands, so that we can more efficiently select and apply the appropriate methods. This should result in more efficient studies, and greater user satisfaction. It should also help move the field of CTA into becoming more of a technology.

INTRODUCTION

Information technologies are affecting more and more work domains. These technologies place higher demands on cognitive skills such as decision making, planning, and maintaining situation awareness. For these reasons, human

Advances in Human Performance and Cognitive Engineering Research, Volume 1, pages 161–199.

ISBN: 0-7623-0748-X

factors professionals are increasingly turning to Cognitive Task Analysis (CTA) methods to design better decision support systems, to design better human-computer interfaces, and to design better training programs. Although several CTA methods are rooted in laboratory techniques for studying cognition, many were developed to address applied needs. The value of CTA methods is to enable the investigator to "get inside-the-head" of the person performing the task, and to understand the strategies and skills that are used.

The goal of this chapter is to suggest some guidelines for conducting CTA projects. We hope that these will be helpful at several different levels. We want to enable the organizations sponsoring CTA studies to determine what to expect from a CTA study, and to estimate the success of a study that has been commissioned. We also hope that these guidelines will be useful for practitioners carrying out CTA studies, to better plan the way that methods are selected and applied. The guidelines we present have emerged from our own CTA projects; we have conducted more than 50 such studies in the past 15 years, and have collaborated on several dozen more (see Klein, 1998, for a description of many of these efforts). In addition, we have learned a great deal from colleagues in other organizations who have performed CTA projects that we consider to be highly successful.

CTA methods may be useful because they offer researchers the opportunity to understand the way people make key judgments and decisions, interpret situations, make perceptual discriminations, solve problems, generate plans, and use their cognitive skills to carry out challenging tasks. When a CTA project helps us appreciate how cognitive skills are used, we can do a better job of supporting these processes, and of removing barriers to them. CTA methods can therefore help practitioners improve the performance of complex tasks that have a strong cognitive component.

Traditional task analysis methods were not explicitly designed to address the subtle, cognitive aspects of task performance. That is why CTA methods can be useful to develop: better training programs, better human-computer interfaces, better decision support systems, better selection procedures, better accident investigations.

Training for tasks that require judgment and decision making is an example of an area of application. Seamster, Redding and Kaempf (1997) point out that cognitive skills require unique methods for assessment as well as for training; training methods that were designed to teach procedural skills do not have the flexibility to handle cognitive skills. The growing use of information technology has resulted in systems that place more cognitive demands on the operators, and the effect is that CTA methods are becoming more important for helping system operators learn how to manage the flow of data and information.

For application in system design, the cognitive engineering approach depends heavily on CTA methods (e.g. Andriole & Adelman, 1993; Essens, Fallesen, McCann, Cannon-Bowers & Dorfel, 1995; Rasmussen, Pejterson & Goodstein, 1994; Woods, 1993). The use of CTA in cognitive engineering studies is one way to factor the users' cognitive requirements into a design.

The increased interest in CTA methods also raises some questions, such as what counts as a CTA study (the boundaries are not very clear), and whether people were examining cognitive task requirements before there was CTA (clearly they were). One of our goals for this chapter is to suggest criteria that might help to reduce these types of confusion. If a research team labels its work as a CTA project, simply because their interviews included two or three general questions about cognition, without getting at the decision strategies employed by the operators, then we might experience some concern about whether this was really a example of CTA. Similarly, an effort that carefully probed the ways operators were interpreting data to arrive at judgments would seem to count as a CTA project, even if the investigators never used that term. The methods that the researchers use to elicit and analyze and represent their data are more important than the labels they attach to the work.

This chapter presents some criteria for what counts as a successful CTA project. We have illustrated these criteria with some examples. We recognize that not everyone will agree with our position, and if that disagreement can turn into informed debate, then that would also be useful.

The organization of the chapter is as follows. The first section presents a brief overview of the progress that CTA has made during the past decade. The second section considers several definitions of CTA. We next present our view of what counts as a successful CTA project. Then we contrast CTA with Behavioral Task Analysis. The next section offers a theoretical perspective for CTA based on aspects of expertise. Then we list some of the most common CTA methods and discuss criteria for selecting CTA methods. Finally, we offer some research directions for further examination of CTA methodology.

PROGRESS IN CTA

In the United States, CTA methods are relatively new. In contrast, European human factors and ergonomics professionals have long been examining cognitive processes. The reason for this divergence is the influence of behaviorism in the U.S., through the work of J. B. Watson and B. F. Skinner. As the dominant paradigm for almost half a century, behaviorism discouraged

U.S. researchers from studying cognition, and the practioners who grew out of this tradition were likewise uncomfortable with examining mental events. The European tradition was more favorable for the study of cognitive processes.

The cognitive revolution in the U.S. took hold during the late 1960s, with the publication of work such as Neisser's book, Cognitive Psychology (1967). The journal of the same name published its initial issue in 1968, and the European journal "Cognition", began in 1972. Although this movement changed the climate for basic research, it did not influence the application of human factors methods.

The field of artificial intelligence created the initial demand to "get inside-the-head" of experts. Artificial Intelligence researchers became active in the period after WWII, and the work of Newell and Simon (e.g. 1972) opened up another perspective on studying cognition, by trying to build computational models of thought. In the late 1970s, Artificial Intelligence researchers put more emphasis on the content of thought, to go along with the algorithms and heuristics that had been studied. The researchers started building expert systems. This work proceeded by eliciting from experts the rules that they followed in performance complex tasks. Hayes-Roth, Waterman and Lenat (1983) presented a comparison of different expert system approaches, to make the technology more accessible to computer scientists. The expert systems framework took Artificial Intelligence out of the laboratory, and into applications.

One of the bottlenecks in building expert systems was in eliciting these rules, in order to fully capture the expertise. The knowledge elicitation proved to be time-consuming and difficult. A new specialty was born, the knowledge engineer, whose job it was to find out how the experts performed their cognitive tasks.

The emergence of knowledge engineering was key to the evolution of CTA. The knowledge engineers created a demand for new and effective methods of getting inside the heads of experts. Because expert systems were intended to replace subject matter experts, it was important to capture the subtle aspects of expertise along with the more obvious rules. The practical goal of expert systems developers was to describe the basis for expertise for specific tasks. This is the same program that we find today for CTA.

The development of expert systems has not flowered in the ways that were expected. One reason may be that rule-based accounts of expertise are too simplistic and too brittle. Today, the leading artificial intelligence proponents have turned to neural nets and genetic algorithms to capture expertise without having to do any knowledge engineering.

However, the legacy of knowledge engineering continued in the field of human factors. Once the methods of knowledge engineering were demonstrated for building expert systems, the applications to other domains were obvious. If it was possible to describe how experts were performing cognitively complex tasks, then these descriptions could also be used for designing intelligent tutoring systems, and for designing training programs. The descriptions could be used in developing new computer systems, and particularly in helping human factors professionals handle the challenges of human-computer interface design. A further impetus to develop CTA methods came from military organizations interested in gaining a better understanding of decision making. The Army Research Institute initiated a program to study decision making in natural settings. In 1989 the Navy followed suit, after the U.S.S Vincennes shootdown of a commercial airliner, and began its program of research on Tactical Decision Making Under Stress (TADMUS), as described by Cannon-Bowers and Salas (1998). These efforts funded additional development of CTA methods. There are several other factors that contributed to the development of CTA methods, including the exploration of cognitive aspects of instruction. Currently, the human factors community is showing an increased level of interest in CTA. A special issue of the journal *Human Factors* (to be edited by Woods and Hoffman) is one example; a special issue of *Ergonomics* is another example. There have been several recent books (Gordon, 1994; Kirwan & Ainsworth, 1993; Klein, 1998; Klein, Orasanu, Calderwood & Zsambok, 1993; Seamster, Redding & Kaempf, 1997; Vicente, 1999; Zsambok & Klein, 1997; Schraagen, Chipman & Shalin, 2000) and articles/reports (Cooke, 1994; Essens et al., 1995) that are partially or entirely focused on CTA. The annual meetings of the Human Factors and Ergonomics Society usually include one or more workshops and sessions on CTA.

The rapid growth of interest and methods offers an opportunity to step back and consider some underlying themes in the field of CTA. This may be a useful time to examine questions such as what constitutes a CTA study.

DEFINITIONS OF COGNITIVE TASK ANALYSIS

To define CTA, it may be useful to first consider the three terms "cognitive," "task," and "analysis." The idea of a *cognitive* investigation is to study the mental processes of the individuals. This does not mean that there are "non-mental" or "non-cognitive" tasks, but rather that the effort is aimed at processes such as decision making, situation awareness, judgment, problem detection, attention management, and so forth. The contrast to a *cognitive* investigation would be an investigation of the steps that a person was supposed to follow in performing a task, and the explicit cues initiating and terminating each step. A *cognitive* investigation would look to see if the cues were all straightforward, or if some were subtle, and others were dependent on context and on the interpretation of the situation. The *cognitive* processes would include the strategies that were used, the skills (such as perceptual discriminations) that had been acquired, and the knowledge (including rich mental models) that was necessary.

The focus on "task" is also important. A CTA investigation is about how people perform *tasks* in natural settings. The *tasks* can be work related, such as generating a weather forecast, predicting the spread of a fire, or noticing signs of infection in a neonate. The *tasks* can fall outside of work, as in shopping in a supermarket or driving. But in all cases, the interest is in how people are performing *tasks* in natural settings, as opposed to basic research tasks.

The third term is "analysis," and this signifies an attempt to extract meaning from the data gathered, to provide an *explanation*. In many situations, researchers can ask subject matter experts questions such as what they were noticing, why they selecting a certain option, how they interpreted the new information. These questions obviously get at cognitive processes for specific tasks. But the questions do not constitute an *analysis*. Similarly, generating a list of important cognitive processes in performing a task (e.g. scanning the data, drawing inferences, storing information) does not constitute an *analysis*. The CTA still needs to explain how these processes are accomplished in the specific domain studied. An *analysis* implies a planful approach for examining the data gathered, identifying cognitive strategies, and representing these as a comprehensive explanation, so that others can learn how the task was done.

Together, the three parts of CTA can be defined as: A set of methods to elicit, explain, and represent the mental processes involved in performing a task. This definition of CTA is reasonably consistent with other definitions that have been offered. Redding (1989) has defined CTA as determining the mental processes and skills needed to perform a task at high proficiency levels, and the changes that occur as skills develop. This seems to be a reasonable definition. It frames the CTA as an investigation into mental processes and skills, and it emphasizes expertise that has either been achieved or is being developed. One quibble is that this definition should include knowledge, along with skills. A second quibble is that many CTA projects do not capture changes during skill acquisition, although these are certainly useful. A third quibble is that CTA is also appropriate when a task is not performed at high proficiency level. CTA is applicable at the novice level (e.g. an investigation of novice drivers (Klein, Vincent & Isaacson, 2001). In addition, it can be productive to contrast the strategies of experts and novices, as a way of learning more about the expertise needed for proficient performance. Nevertheless, for most applications, CTA is directed at trying to understand the cognitive skills, knowledge, and strategies needed to achieve high proficiency levels. In addition, CTA studies often try to identify the barriers to attaining or applying cognitive skills. Thus, in helping to design training programs, human-computer interfaces, and decision support systems, the intent is to help people develop and exercise expertise.

Essens et al. (1995) describe the *product* of a CTA as "a task-specific cognitive model including hypotheses about cognitive performance" (p. 43). This is consistent with Redding's definition. It means that the CTA needs to result in an understanding of how a person is thinking and judging and sizing up situations. However, this definition raises the question of what counts as a model, a topic that can sometimes be contentious. The concept of a model in this definition seems to correspond to the concept of an explanation in our definition.

Militello, Hutton, Pliske, Knight and Klein (1997) have defined CTA as the description of the cognitive skills needed to perform a task proficiently. This follows Redding's concern with capturing expertise, and with clarifying the value added of experience.

Roth, Woods and Pople (1992) state that "A Cognitive Task Analysis depends on two mutually-reinforcing activities: an analysis of the cognitive demands imposed by the world that any intelligent agent would have to deal with (a model of the cognitive environment); and an empirical investigation of how practitioners, both experts and less-skilled individuals, respond to the task demands (performance models)" (p. 1195). This extends Redding's framework, to describe the challenges posed by the world in which a person is acting. To be useful, a CTA must connect with the constraints of the world, particularly constraints in the task environment. Some researchers (e.g. Vicente, 1999) argue that the exploration of these constraints is itself a useful form of analysis, and perhaps even a starting point for the cognitive exploration. We agree with Roth et al. and with Vicente about the importance of tying the cognitive analysis to the constraints of the situation. The importance is reflected in the term "task," because the investigation must capture the situational constraints in specifying the task that is being studied.

Gordon and Gill (1997) have claimed that CTA is most suited for cognitively complex tasks with an extensive knowledge base, complex inferences and judgment, in a complex, dynamic, uncertain, real-time environment. We agree. However, we have found that it is possible to perform effective CTA studies even when the tasks are not dynamic, or highly complex, as in studies of the strategies used by consumers. Designers of consumer products need to understand consumer decision making regarding the purchase and use of products. This understanding is critical and very difficult to obtain without CTA methods. These CTA studies may not be as dramatic as studies of pilots, but they can still be informative and useful.

One way to clarify the definition of CTA is to provide counter-examples, describing what CTA is *not*. These counter-examples include:

- basic research into cognition
- a prescription about how people should be thinking
- a behavioral task analysis
- a list or diagram of cognitive tasks that could be added to a curriculum.

CTA is not research into the nature of basic cognitive processes. Certainly, a basic research investigation could include CTA methods; our assertion is that basic research into cognition does not automatically qualify as a CTA study. Because of the difficulty of setting up controlled conditions, few studies of basic cognitive processes investigate the types of tasks performed in natural settings. Traditional cognitive research does not attempt to determine the cognitive demands imposed by the world (whether the world of a pilot, a nuclear power plant operator, a nurse, or a supermarket shopper) and does not attempt to empirically determine the way people respond to these demands. We do not imply that traditional laboratory studies are unimportant. On the contrary, the insights provided by basic research about cognitive processes such as attention, working memory, categorization, and decision making can be valuable for developing better CTA methods and for interpreting results. Moreover, some basic research programs (e.g. Meyer & Kieras, 1997) have important implications for applied projects. And the findings that emerge from such research can be incorporated into CTA efforts. However, our claim is simply that basic cognitive research studies that do not investigate how people perform natural tasks are not examples of CTA.

One way to think about this distinction is that there is a difference in the goals of CTA projects and basic research studies. CTA projects attempt to explain phenomena. These phenomena, such as how a military pilot adjusts to unexpected enemy air defenses, how a shopper selects a product, how a baggage screener notices a problem in just a few seconds, cannot be explained in terms of steps or rules to be followed. There is more to the story, and that is the part that CTA seeks to fill. In contrast, basic research attempts to test theories. Even when basic researchers study natural settings, their focus is on the theories, and the hypotheses drawn from these theories.

CTA is not a prescription about how people *should* be thinking, using normative theory. The focus of CTA is on learning what people actually do, what they actually know. This is very different from analyzing what they should be doing and setting up a condition to demonstrate that they did not follow an optimal procedure. While studies focusing on optimal procedures can sometimes be informative, often they depend on being able to eliminate complicating contextual factors. Optimization becomes very difficult to measure in natural contexts (e.g. Klein, 2000). But once the context is limited, the basis for expertise is also limited. In contrast, the goal of a CTA project is to learn more about expertise, and particularly to discover facets of cognitive skill that had not previously been recognized. It is only after the cognitive processes of proficient performers are understood that CTA researchers have a basis for suggesting how decisions and judgments should be made.

CTA is different from behavioral task analysis. This distinction has been mentioned earlier, and will be discussed in greater detail in a later section.

The output of a CTA study should go beyond a list or diagram of cognitive tasks. The listing of tasks does not constitute an analysis or explanation. One concern here is that the researchers will generate a list or diagram of cognitive tasks as if these tasks should be added to the existing training requirements. When this happens, one of the worst features of behavioral task analysis (a non-integrated listing of tasks) is made even less palatable by piling cognitive tasks onto behavioral ones.

CRITERIA FOR SUCCESS IN CTA PROJECTS

We can distinguish three criteria for success of a CTA effort. A project needs to make an important *discovery*, it needs to *communicate* this discovery, and the communication needs to result in a meaningful *impact*.

Discovery

What counts as an important discovery? For a CTA project to be counted as successful, this means that the CTA team has learned something new and important about key judgments and decisions. The CTA team has identified judgments and decisions and other cognitive demands about which the sponsors or potential users were unaware, the strategies used to make the decisions, or the patterns used to make the judgments. The discoveries can be about strategies of making decisions, as in the determination that highway engineers use different strategies to make different types of decisions (Hammond, Hamm, Grassia & Pearson, 1987). The discoveries can be about patterns of cues, as in CTA efforts to define how nurses detect sepsis in infants (Crandall & Getchell-Reiter, 1993), or how nurses detect necrotizing enterocolitis (Militello & Lim, 1995). Other examples include identifying the information-seeking strategies used by skilled weather forecasters (Pliske, Klinger, Hutton, Crandall, Knight & Klein, 1997); describing the decision strategies of fireground commanders (Klein, Calderwood & Clinton-Cirocco, 1986); discovering how milk delivery personnel are able to rapidly and accurately prepare the right number of containers (Scribner, 1985); determining how Micronesians are able to navigate long distances without navigational equipment (Hutchins, 1983); and describing how anesthesiologists adapt to the new cognitive and physical demands created by the introduction of a physiological monitoring system (Cook & Woods, 1996).

Figure 1 shows a range of outcomes that could emerge from a CTA study, from the enumeration of cognitive functions, through the dissection of these functions, and up to the generation of an explanation of how the function is accomplished. At a minimum, a CTA investigation should identify the cognitive functions that are required. This would include describing the types of judgments and decisions that are needed. However, just listing these types of judgments is usually not enough. The potential users are likely to examine the list and comment that there is nothing on it that is surprising, or useful.

Figure 2 presents an example, taken from the domain of aircraft operation. The key types of decisions are presented, but they are not unpacked. We see where the tough decisions are, but we can only guess about the type of expertise needed to make each of them. This should be the starting point for the CTA study, not the conclusion. In Fig. 2, further analysis might show that

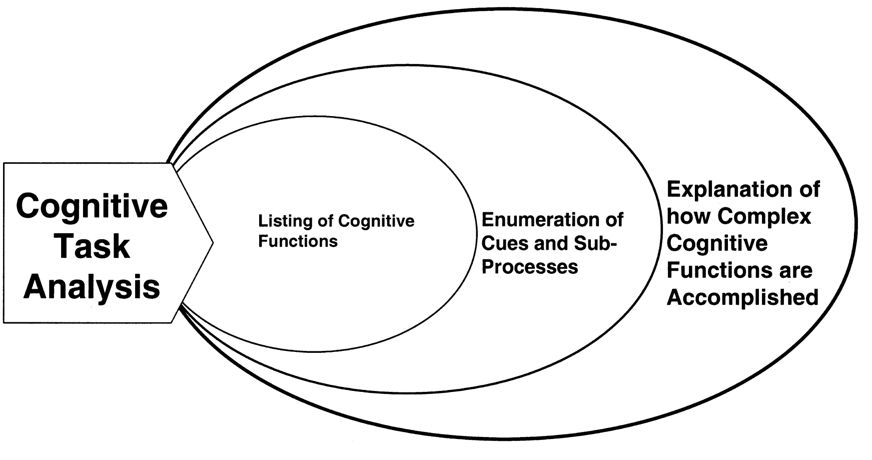

Fig. 1. Range of Outcomes for CTA Projects.

2.1.1 Perform normal takeoff operations

2.1.2 Perform takeoff roll procedures

- **2.1.2.1 Position aircraft on runway centerline and stop**
- **2.1.2.2 If required, transfer control of aircraft**
- **2.1.2.3 If required, comply with standard policy for transfer of aircraft control**
- **2.1.2.4 Select HDG HLD on MCP**
- **2.1.2.5 Set WX radar for takeoff**
- **2.1.2.6 Release brakes and set takeoff thrust**
- **2.1.2.7 Advance power on both engines**
- **2.1.2.8 PF call for EPR**
- **2.1.2.9 PNF select EPR on MCP**
- **2.1.2.10 Engage auto-throttle in EPR mode as engines are accelerating through 1.1 EPR**
- **2.1.2.11 Comply with standard policy for takeoff and go/no go decision**
- **2.1.2.12 Maintain directional control**
- **2.1.2.13 Monitor engine and flight instruments**
  - **13.1 If abnormality exists, Captain decides and initiates Rejected Takeoff**
- **2.1.2.14 Complete standard callouts for takeoff conditions**
- **2.1.2.15 Comply with Standard Policy for Takeoff Flight Path Control Techniques**
  - **15.1 Recognize unstable flight path condition and be prepared to execute immediate recovery**

Fig. 2. Portion of Task Analysis for Aircraft Takeoff.

several of these functions can be performed more effectively by training on a few skills, such as building up a better sense of typicality.

Returning to Fig. 1, the next level of discovery is to describe the cues and patterns and strategies that go into the judgments. Table 1 presents a list of cues taken from Crandall and Getchell-Reiter (1993) who studied nurses in neonatal intensive care units. These are the cues that the nurses had learned to recognize as possible precursors of sepsis. At the time of the research, approximately half of these cues had not appeared in the medical or nursing literature. Therefore, the cataloging of cues was an important discovery in its own right. The nurses initially had believed that they could not generate any explanations for how they were making an early detection of sepsis, and it took hard work, using the Critical Decision Method to probe specific incident accounts, in order to obtain this list.

The most exciting discoveries in a CTA study are those that result in an explanation or insight regarding the way a cognitive function is performed. Most CTA studies do not achieve this outcome, but we believe it should be established as a criterion for success. Thus, Smith, Ford and Kozlowski (1997) used simulated malfunctions to study pilot reactions, and showed in detail the differences in mental models that resulted in effective versus ineffective troubleshooting. Cook, Woods, Walters and Cristoffersen (1996) studied medical evacuation personnel and went beyond the surface features of the task to describe the underlying decisions regarding tradeoffs between resources; this resulted in a better design for a decision support system. Crandall and Getchell-Reiter (1993) did not leave their analysis at the listing shown in Table 1. They described how nurses used patterns of cues. The detailing of these nursing decisions was arranged in the context of specific incidents that could be used for training. Klein et al. (1986) described the decision-making strategies of firefighters as a recognition-primed decision model.

Studies such as these demonstrate how a CTA study can result in an explanation for a phenomenon that previously was not well understood. These

Table 1. Indicators of Sepsis.

Indicator	Identified in medical literature		Number of cases in which expert nurses checked for indicator
	Yes	No	
Unstable temperature	X		3
Hypothermia	X		2
Elevated temperature	X		
Feeding abnormality	X		4
Abdominal distention	X		5
Increased residual	X		2
Vomiting	X		
Lethargy	X		8
Irritability	X		
Color changes	X		9
Respiratory distress	X		4
Apnea	X		6
Cyanosis	X		2
Tachypnea	X		
Seizures	X		
Jaundice	X		
Purpura	X		
Bradycardia		X	7
Unresponsive		X	5
Poor muscle tone		X	5
Perfusion or mottling		X	2
Edema		X	2
“Sick” eyes		X	2
Clotting problems		X	2

explanations can be valuable for the sponsors and potential users of the research, and the findings often will contribute to our basic understanding of expertise and decision making in field settings.

Communication

What counts as effective communication of the discovery? Representation of findings can be more difficult than knowledge elicitation. Successful communication occurs when the CTA team provides the potential user of the findings with an understanding that is sufficiently vivid so that the user can take the cognitive processes into account in designing an intervention. The CTA representation is attempting to generate the response, "Now I see how they are really doing that job." If the project is a collaborative effort, and the person who performed the CTA is part of a design team, she/he can represent the user's perspective during the design process, and representation may not be as important. However, if the CTA results are handed off, then it becomes critical for the recipients (i.e. system developers, instructional designers, etc.) to appreciate the operators' cognitive skills. In many cases, operators are themselves unable to articulate their own strategies (e.g. Nisbett & Wilson, 1977) and the designers of systems or training may not be skilled at inferring these strategies. Therefore, the role of the CTA is essential as an advocate for the subtle cognitive skills that must be supported if the task is to be performed effectively.

An example of effective communication comes from some work by Cook et al. (1996), in their study of the process of medical evacuation in military operations. Cook et al. went beyond the existing list of tasks and functions, and portrayed the medical evacuation personnel using a problem space. Inside that space was a region in which the resources (for slots on aircraft) were greater than the demand. For that region, there was little benefit for trying to optimize the solutions. In a second region of the problem space, the resources were less than demands. Here, the job was less about configuration and prioritizing than it was about negotiating for more resources, or for permission to postpone service. The third region was marked by uncertainty about whether the resources were sufficient to meet demand. Here, the system operators would need help in finding efficient configurations. The immediate task for the operators was to determine which region of the problem space they were in. This approach to representation was very useful for the system designers to help them understand the needs of the operators.

Impact

What counts as having an impact? When the CTA findings are put into action, this criterion is achieved. CTA projects are initiated for a purpose, and the intended users of the findings usually know how they want to apply the results. The goal of the study may be to help the users change the way they design a product or a system, or to revise the way they are training people, or even to confirm the users' current practices so that they can press on rather than continuing to gather more data. Sponsors may find uses for basic research types of CTA studies seeking to identify strategies of decision making, problem solving, planning, or situation awareness.

In the domain of air traffic control (Redding & Seamster, 1994) a comprehensive CTA of aviation training was conducted, resulting in a redesign of the Federal Aviation Administration's controller training curriculum. In the domain of avionics troubleshooting (Gott, Pokorny, Kane, Alley & Dibble, 1996) an extensive CTA of expert troubleshooting skill was conducted. The U.S. Air Force is in the process of developing an intelligent tutor based on this CTA. A second system developed along these lines (Pokorny, personal communication, 1997) is an expert system for supplementing personnel policy administrators in the U.S. Air Force. The resulting system is currently in beta test.

It is generally difficult to measure the eventual value of an intervention. Evaluation of impact generally involves gathering longitudinal data and taking into account intervening variables, neither of which are trivial tasks. For most CTA projects, funding ends at the time the intervention is introduced. Nevertheless, it is possible to obtain some data on the impact of the intervention that was based on CTA findings.

Another example of impact comes from the work of Di Bello and her colleague (e.g. Di Bello, 1997; Di Bello & Spender, 1996). Their work was with maintenance technicians in the transportation industry, to help improve the use of new information technology tools. Most of these tools, for scheduling, workflow, spares management, and so forth, are rejected in the workplace. A primary reason for this rejection is that the workers themselves do not have confidence in the tools, or understand how to apply them. Di Bello conducted a CTA study to understand how bus mechanics were scheduling their work. The CTA method consisted of observation and interview using two simulations. The first simulation presented a mechanic with a set of actual work orders and requested that the mechanic review these and generate a schedule for performing repairs. The second simulation presented a mechanic with longitudinal data on a single bus, to see what inferences the mechanic could derive. Di Bello and Spender found that the less effective mechanics lacked a mental model for seeing patterns in the longitudinal data, and were therefore forced into a reactive profile (repairing what came in) as opposed to a proactive stance to anticipate problems and prevent them. These findings were translated into a training program using a fleet of 30 buses in a low-fidelity simulation to enable the mechanics to see the limitations of their reactive strategy, and to realize the value of anticipating problems. With their shift in mental models the mechanics were able to appreciate how the new software scheduling program provided them with the data they needed. Evaluation data showed that prior to the training, 70% of the mechanics were categorized as showing a reactive mental model (versus 30% who showed a proactive mental model). After the training, the ratio had reversed, and 70% of the mechanics showed a proactive mental model for anticipating problems and preventing them.

Klinger and Militello (in preparation) modified the human-computer interface for Airborne Warning and Control System (AWACS) weapons directors. The CTA study showed how the weapons directors were losing situation awareness because of memory limitations. Each time the weapons directors looked away from the screen (to operate switches), they needed time to re-orient because all the icons on the screen had moved. In addition, the operators had difficulty keeping track of key icons such as high threats or assets such as tankers for aerial refueling. The human-computer interface solutions were straightforward (marking the key icons, using an on-screen menu to double for the most common switch actions). A laboratory study using high-fidelity simulators indicated that the redesigned interface resulted in a significant improvement in performance. In addition, the redesigned interface was selected to be incorporated in the next modification cycle for the AWACS weapons directors' workstations.

Andriole and Adelman (1993) provide some additional examples of cognitive engineering projects that resulted in improved performance. These studies show that it is possible to measure impact, but this is done most easily during the development of pilot programs or prototypes.

What does *not* count as a successful CTA study? Often reports of CTA projects include pages of tables and matrices reporting data in exhaustive detail, demonstrating the amount of effort expended, but these reports do not communicate discoveries. If CTA methods are to be judged useful, they must produce discoveries that are useful to the sponsoring organization. Careful data collection and exhaustive reporting are not a substitute for learning and describing something new and important about the way specialists think about key tasks. We recognize that there are projects in which a detailed reporting of data is required, or in which pithy summaries of discoveries are not possible, but the criterion for success can be whether or not the findings were used to help the sponsoring organization make a change, and to describe how the findings contributed to the change.

COGNITIVE TASK ANALYSES VERSUS TRADITIONAL TASK ANALYSIS

We can distinguish between CTA and more traditional task analysis methods that focus on behaviors. (We will refer to the latter as behavioral task analysis.) CTA studies address the cognitive requirements for a task, whereas behavioral task analyses focus on the observable behaviors and cues. It is possible to perform a behavioral task analysis, to develop task sequence diagrams, and to specify the initiating and terminating conditions for sub-tasks, without considering cognitive issues of how the cues are recognized, how the sub-tasks are interpreted, or how the situational interpretation affects the strategies selected. One of the strengths of behavioral task analysis is its emphasis on the objective nature of the data: explicit tasks and sub-tasks, specific and unambiguous cues, and clear criteria for terminating one step and initiating the next. These advantages are achieved by steering clear of the ambiguous, context-bound issues that are central to CTA.

Kirwan and Ainsworth (1993) do not present a sharp distinction between CTA and behavioral task analysis. They define task analysis as "the study of what an operator (or team of operators) is required to do, in terms of actions and/or cognitive processes, to achieve a system goal. Task analysis methods can also document the information and control facilities used to carry out the task" (p. 1). Kirwan and Ainsworth describe more than 40 different task analysis approaches, including several CTA methods. We have chosen to accentuate the distinction between CTA and behavioral task analysis because of our concerns that a cognition-free behavioral task analysis may provide a misleading account.

Some would argue that skilled task analysts are already doing many of the things that are discussed in the CTA literature, so CTA does not offer anything new. We agree that there are many experienced practitioners of behavioral task analysis whose practices include ways of probing subject matter experts to capture different types of cognitive skills. Therefore, behavioral task analysis may not preclude CTA. Nevertheless, we see a distinction. Traditional task analysis usually does not call for probes of cognitive processes, and rarely provides guidance in conducting such probes. Hence, the methods themselves do not necessarily lead to the capture of the cognitive aspects of skilled performance. CTA may be seen as complementary to traditional task analysis

– a way to enrich the accounts of how the steps of a task are performed. But the two, CTA and behavioral task analysis, are different. CTA methods provide tools for investigating cognitive elements in the context of real-world constraints and affordances of the tasks.

In our experience, CTA efforts can benefit from having previous behavioral task analysis data to describe explicit functions that need to be performed. Similarly, we believe that many behavioral task analyses can benefit from having CTA data to clarify how the work is actually being accomplished. Otherwise, there is a risk that behavioral task analyses may distort the picture of performance. Traditional task analysis seeks to decompose tasks into smaller elements, and to generate orderly hierarchical accounts.

The way in which the Instructional Systems Development process has been traditionally applied is a prime example of such an approach. Although the process has been successfully applied in describing highly proceduralized tasks, the resulting representation can oversimplify cognitively loaded tasks. This strategy of hierarchical decomposition does not provide a systematic means to capture and explore the important cognitive skills that may underlie tasks and subtasks identified. Often the resulting representation is a series of steps with no explanation of the cognitive strategies needed to accomplish the steps satisfactorily.

The interpretation of a task list is that if one follows the steps identified, the task should be accomplished. However, for cognitively loaded tasks, subject matter experts often do not follow those steps, and are not following any steps at all (Klein, 1978). In these cases, traditional task analysis may result in a partial and potentially misleading account of performance. The requirement in an Instructional System Development project to write testable objectives for skills generates further pressure to decompose a task into small steps that can easily be checked off or assessed via multiple choice questions. The steps may have little psychological reality, but are a means to permit objective evaluation. When this goes too far, the so-called steps of the task are not accurate descriptions of proficient performance. Here, CTA can provide a more accurate account of how specialists move beyond the rule-based level of performing complex tasks.

Thus, we argue that CTA studies have the challenge of describing the decision-making, judgment, and problem-solving strategies used to perform tasks. The descriptions can elaborate on existing task descriptions. DeMaio, Parkinson, Leshowitz, Crosby and Thorpe (1976) studied instructor pilots who were taught a rigid procedure for visually scanning flight instruments. Eye movement data revealed that the instructor pilots did not themselves follow the procedure, and were better able to detect anomalies than the students. One risk of behavioral task analyses is that they can promote the rigid adherence to inefficient practices, such as over-learning a scanning strategy, rather than developing the associations and patterns that would permit rapid attention to the flight instruments that were most relevant. One of the promises of CTA is to allow us to break loose from the official procedures, and discover the bases for expertise within a domain.

FRAMING CTA EFFORTS AROUND ASPECTS OF EXPERTISE

At some point in the future, we may be able to work from a theoretical account that would guide the application of CTA methods. The availability of an overarching perspective would help to develop CTA into a technology: determining the boundary conditions of different CTA methods, evolving guidelines for selecting CTA methods to match the features of the situation, deriving formats for reporting findings, establishing metrics for evaluating the degree of success of the CTA effort, preparing training programs for using CTA methods, and setting standards for practice.

We are not close to that point, and we may find that it is not necessary. Behavioral task analysis methods have proliferated and have been applied without a theory of the task. Practices in designing human-computer interfaces have improved without theory. The maturation of practice can be sufficient to achieve progress, without requiring a top-down direction.

Moreover, the prospects for theory are not great. A theoretical account might have to cover the nature of expertise, the functional distinctions between types of tasks, constraints in the world of practice, and the features of the cognitive processes themselves. We are not close to having a comprehensive theory in any one of these areas, let alone an integration of all of them.

An additional distinction can be made regarding the assumptions behind the method. Thus, some methods have emerged from rule-based information processing accounts of cognition, whereas others are linked to phenomenological descriptions of cognition, still others can be traced to ethnographic techniques, and others to systems engineering approaches. Therefore, the conceptual background regarding the field of psychology may have to be taken into account.

Therefore, the ambition of deriving a taxonomy of CTA methods may be premature. Attempts to develop such a taxonomy have faltered, as the analysts found that the initial distinctions and characteristics of CTA methods carried little force for directing which methods should be used under what conditions. We can now see that a taxonomy would require that we know how to characterize tasks and situations, in order to recommend which technique to use for what type of project. A taxonomy would require the theoretical accounts listed at the beginning of this section.

Despite these problems, we do not claim that efforts at theory building should be suspended until all of the components are firmly in place. The attempt to develop theory, and to derive taxonomies, allows us to mark our understanding of CTA methods, and to make this understanding explicit to ourselves and to others. Moreover, the attempt to formulate theory regarding CTA may help to clarify what is needed from a theory of expertise, or a theory of cognitive processes.

To attempt some contribution to theory building, we will describe our current perspective on the aspects of expertise that practitioners try to capture using different CTA methods. By examining some standard sources on cognition (Barsalou, 1985) and expertise (Chi, Glaser & Farr, 1988; Ericsson, 1996; Ericsson & Smith, 1991), we have compiled a sample of important aspects of skilled performance of complex tasks. We have also realized that it may be useful to distinguish between aspects of expertise that reflect what a person knows, and the ways that people can apply the knowledge. This distinction is presented in Table 2.

Our belief, based on a literature review and on experience conducting CTA projects, is that we can identify a small set of different types of knowledge that experts have. They have a set of associations (e.g. Sloman, 1996, on associative inferences versus rule-based inferences; also see Simon, 1956, on pattern learning) which include the judgment of what is *typical* in a situation. And this type of knowledge permits experts to quickly spot anomalies that are departures from typicality. Experts have formed *mental models* which are usually causal frameworks explaining how things happen. And this type of knowledge permits experts to mentally simulate what will happen in the future, based on the current configuration of events and states. It also permits experts to discover workarounds when a plan is blocked, using their mental models of how to introduce changes to lead to a desired effect. Experts have learned a great many *routines* for getting things done, and this permits them to make rapid decisions. Experts have learned to make *perceptual discriminations*. And this enables them to see things that novices cannot detect. Finally, experts have learned a great deal of *declarative* information. However, CTA studies usually do not try to capture this type of knowledge because it is so vast, and because it is fairly accessible without any special CTA techniques. Therefore, declarative knowledge is not included in Table 2.

The match between skills and knowledge is obviously more complex than shown in Table 2. Thus, the ability that experts have to make perceptual

Table 2. Aspects of Expertise to be Captured by CTA Studies.

Knowledge: What you have to know	*Skills*: What you can do with that knowledge
Typicality (patterns)	Spot anomalies
Mental models (causal dynamics)	Assess situations
Routines	Perform workarounds
Perceptual discriminations	Mental simulation of past and future (expectancies, time horizons, prioritizations, attention management, and synthesizing information)
	Achieve rapid decision making, detect affordances in a situation, apply heuristics, and detect problems

discriminations can help them spot anomalies and also detect early signs of problems. We consider the breakout in Table 2 to be a potentially useful step for preparing to do a CTA study. It may be useful to consider what types of knowledge and cognitive skills the CTA should emphasize, and even to reflect on the CTA approach that might be well suited for that type of knowledge or skill.

Some tasks depend on the ability to make fine discriminations between similar cues and patterns. Shanteau (1992), Klein and Hoffman (1993), and others have described how specialists have better *perceptual skills* in a domain than novices. We suggest that perceptual discriminations are challenging in a CTA study because they are difficult for subject matter experts (SMEs) to describe. However, by having subject matter experts relate an actual incident, the cues and discriminations may be more easily accessed from memory.

Some tasks depend heavily on being able to understand the causes and interactions in a situation. Specialists appear to use their experiences and feedback to build up a richer map of the causal connections. Therefore, to capture the basis of their expertise, we need to clarify the nature of their *mental models* (for a basic overview of mental models, see Rouse and Morris (1986)). However, mental models are notoriously difficult for SMEs to describe. One strategy that has worked for us is to use a simulation that requires the SME to perform a workaround, because in this context the causal influences emerge more easily.

One reason we prefer to use an incident-based method such as the Critical Decision Method is that it seems well-suited for capturing perceptual discriminations, and also because the incidents can be probed in a way that requires workarounds ("but what would you have done if . . ."), thereby reflecting aspects of the SME's mental model. In addition, the incidents also describe some of the skills that are based on expert knowledge (e.g. incidents involving rapid decisions, or early problem detection).

Some tasks are dynamic and time pressured and depend on quick reactions to fleeting events. Specialists are able to control their attentional capacities more skillfully than novices are (e.g. Gopher, Weil, Bareket & Caspi, 1988; Hanish, Kramer & Hulin, 1991). For dynamic tasks, it may be necessary to learn how this *attentional control* is developed and manifested.

Specialists often need to learn *typicality*, in the form of domain-specific prototypes, so that they can recognize typicality, and with that, can recognize anomalies (see the work of Logan, 1988).

Some tasks require people to *synthesize*, to go beyond the evidence given, to draw out inferences and make projections. Specialists appear to have available more heuristics, not in the technical sense of strategies for searching through problem spaces, but in the general sense of having more informal rules and more "tricks of the trade." They know shortcuts and context-specific rules of thumb and *routines* that are often not documented. They can rely on mental simulations (Klein & Crandall, 1995) to project trends and to build explanations.

Specialists may not have "better" memories, as shown by research on chess players (e.g. de Groot, 1946/78, 1986; Holding, 1985). Their strength is in being able to use better strategies for the management of working memory and for long-term storage and retrieval (e.g. Ericsson, 1996). Therefore, we may want to study their strategies for *memory management*. The control of working memory appears related to workload management. The control of long-term memory is important because most domains require specialists to acquire a great deal of declarative knowledge. Therefore, it can be useful to study how specialists obtain this knowledge, how they organize it, and how they retrieve it from long-term memory.

Many tasks require people to monitor themselves and adjust their performance in line with perceived difficulties. The literature on *metacognition* documents important skill development (Means, Salas, Crandall & Jacobs, 1993; Metcalfe & Shimamura, 1994; Smith et al., 1997). It can be useful to find out what strategies experts are employing in order to gauge how well they are doing, and how to adjust to conditions.

If we can use CTA to understand the way competent performers employ these cognitive processes, and particularly if we can determine the way they differ from novices and from mediocre performers, then the effort can have important benefits. The goal of a CTA is to help us to move beyond the behaviors, and into the mental activities of the skilled worker. The goal is also to teach us what the person with expertise is really thinking about. Table 2 simply lists some of the types of knowledge and skills that contribute to expertise. A useful explanation would take the CTA findings and clarify how SMEs perform a task. We discussed this earlier when we examined the nature of discoveries emerging from CTA projects.

Doubtless, the list of aspects of expertise that is presented in Table 2 can be expanded and reorganized. Our intent is *not* to provide a definitive account of the cognitive facets of expertise. We are simply trying to show that there *are* different facets, and that an effort to study the cognitive aspects of task performance can draw on these and other processes. If a task does not present any of these types of challenges, then there is probably little reason to conduct a CTA. For example, a study (Kaempf, Thordsen & Klein, 1991) of Army soldiers in charge of setting up communications gear did not generate any discoveries about their strategies or tacit knowledge. We believed that they did not have to acquire any "real" expertise, because the task was so heavily proceduralized. The soldiers claimed that what marked the real experts was being able to recall the exact page in the manual that documented the steps to be followed.

We should note that the specification of different aspects of expertise, shown in Table 2, should not be taken as a type of conclusion from a CTA study (e.g. this task requires mental models, and judgment of typicality, but not perceptual skills). Rather, the CTA study would need to fill in the depth about the structure of the mental models, the types of prototypes that are recognized, and so forth. Table 2 is a suggestion for the starting point of a CTA study, not the outcome of such a study.

DIVERSITY OF CTA METHODS

Within the past 10–15 years, a wide variety of CTA methods have emerged. Up to this point, we have been discussing CTA at an abstract level. Here, we want to get more concrete about the specific CTA methods that are available. These methods have been described in several different sources (Cooke, 1994; Crandall & Getchell-Reiter, 1993; Crandall, Klein, Militello & Wolf, 1994; Essens et al., 1995; Gordon, 1994; Gordon & Gill, 1997; Hoffman, Shadbolt, Burton & Klein, 1995; Kirwan & Ainsworth, 1993; Redding & Seamster, 1994; Schraagen, Chipman & Shalin, 2000; Vicente, 1999).

Table 3 sorts out the different methods for knowledge *elicitation*, getting the information "out of the heads" of the people performing the tasks. We have found it convenient to sort these into four categories: (a) methods that use interviews (both high and low in structure), (b) observations such as keystrokes

Table 3. Knowledge Elicitation Methods.

Interview Methods	Observation Methods	Modeling Methods	Experimental Methods
Unstructured interviews	Direct observation-questioning	MIDAS	Multi-dimensional & other types of scaling
FAST	PARI	COGNET	Domain knowledge tests
Concept maps	Process tracing	SOAR	Secondary tasks
Conceptual graph analysis	Simulation- High/Low Fidelity Freeze frames	ACT	Bruswik Lens model
Critical decision analysis	Constrained processing- Missing information Time pressure		
Constructed scenarios	Expert/Novice comparisons		
Repertory grid			
Knowledge audit			

(Kieras, 1988; Zachary, Zaklad, Hicinbotham, Ryder, Purcell & Wherry, 1991), and simulations (Woods, 1993; Gott et al., 1996), (c) computer models, and (d) laboratory tasks. Hoffman et al. (1995) have adopted a similar categorization, looking at the analysis of familiar tasks, unstructured interviews, structured interviews, and contrived tasks.) Some CTA methods focus on studies of specific incidents (e.g. Critical Decision Method, Klein et al., 1986), whereas others elicit more general information (concept maps, McFarren, 1987; conceptual graph analyses, Gordon, Schmierer & Gill, 1993; Knowledge Audit, Militello & Hutton, 1998). Hoffman et al. (1995) and Cooke (1994) have discussed the knowledge elicitation strategies in common use.

Table 4 presents different methods for knowledge *representation*, dividing these into methods that describe specific incidents, and methods that synthesize incidents (or do not work from specific incidents). Some of the methods provide representations of specific incidents. Woods, O'Brien and Hanes (1987) describe accounts of nuclear power plant control room crew members reacting to simulated accidents. Kaempf, Klein, Thordsen and Wolf (1996) present records of situation awareness shifts during navy anti-air operations.

Table 4. Knowledge Representation Methods.

Incident-Based Methods	Generic Methods
Problem Behavior Graphs	Expert systems
Decision Flow Diagrams	COGNET
Discourse Analysis	Decision Requirements Table
Annotated Interviews	Concept Maps
Expert/Novice Contrasts	Conceptual Graph Analysis
	Critical Cue Inventory
	Knowledge Audit
	Influence Diagram
	Prototype Descriptions

Other methods provide representations of generic knowledge structures. For example, conceptual graphs (Gordon et al., 1993) provide insights into a person's mental model of a domain. Militello (2001) has described knowledge representation methods in greater detail.

CRITERIA FOR SELECTING CTA METHODS

What is the best method for doing a CTA? It should be obvious that this is a good example of a poor question. In selecting a CTA method, we need to consider the types of cognitive processes to be understood, which are a function of the type of task that is being investigated. A study of pilots flying sorties and performing the task of target acquisition needs to examine the way they control and shift their attention, and the way they perceive subtle changes in their situation. A study of librarians performing the task of recommending books to patrons, needs to examine the librarians' mental models, and can minimize the assessment of perceptual skills. (Perceptual skills might be important for a librarian performing the task of identifying books that are too badly worn to remain in circulation.)

In selecting a CTA method we also need to take into account the way the results are going to be used. If the results are going to be used to determine whether or not the job of an operator can be automated, the study has to describe all of the subtle aspects of expertise that the operator has developed, because these might no longer be available. If the results are going to be used to design a decision support system, the study has to clarify the way the operators are currently making critical and difficult decisions, so that the system will help rather than interfere with them. If the results are going to be used to build training, the CTA must uncover both high-level strategies and concepts to frame the training, as well as specific critical cues, perceptual discriminations, and contextual elements needed to make cognition visible to learners.

Some CTA methods seem to do a better job of capturing certain types of cognitive processes, such as mental models, and others appear more sensitive to different processes, such as perceptual skills, metacognitive strategies, and attentional control. In short, there is no one "best" method of CTA.

If we had a full taxonomy of CTA methods, we would be able to do a better job of matching the method to the types of expertise to be elicited, and the types of applications for the findings. As an initial step, we will present some of our beliefs about the matches between the aspects of expertise covered in Table 2, and the CTA strategies for knowledge elicitation offered in Table 3.

Mental Models

Many CTA methods concentrate on eliciting mental models – declarative knowledge and how it is organized. Concept maps (McFarren, 1987) and conceptual graph analysis (Gordon et al., 1993) fit here; both are ways of having people diagram the primary factors involved in understanding a domain, along with specifying clear linkages between the different types of concepts. Methods for multidimensional scaling (Cooke, Durso & Schwanaveldt, 1994) are another approach for eliciting mental models, by summarizing subjects' responses to paired comparisons. Layton, Smith & McCoy (1994) demonstrated the use of process tracing to describe how the level of completeness of a mental model determined a pilot's troubleshooting success in handling a malfunction.

Attention

Few CTA methods address the way people manage attentional resources. This is probably most important for dynamic, time-pressured tasks, and may be best studied using moderate to high fidelity simulations. Another way to study management of attentional resources is by using devices, such as eye movement trackers, during actual task performance. We have not found cases where eye movement trackers have been incorporated into Cognitive Task Analyses.

Perceptual Skills

These are extremely important in natural settings. The Critical Decision Method (Hoffman, Crandall & Shadbolt, 1998) and the Knowledge Audit (Militello & Hutton, 1998) are two methods for capturing perceptual skills. The Critical Decision Method does this by using a semi-structured interview to study challenging incidents, and to probe about the cues and patterns that were noticed during these incidents, particularly those that might have been missed by inexperienced personnel. The Knowledge Audit is a streamlined method for getting an overview of the types of skills, including perceptual skills, needed to perform the task.

Recognition of Typicality

The essential issue here is to identify anomalies as well as prototypes. In some natural settings, the ability to quickly identify an anomaly, in order to identify a potential problem, is important (Hoffman, 1987). It would seem that the experimental techniques for studying the way people form prototypes and categories (e.g. Kahneman & Miller, 1986; Logan, 1988) could be adapted here.

Routines and Strategies

Lesgold, Rubinson, Feltovich, Glaser, Klopfer and Wang (1988) have described the use of the Precursor, Action, Result, and Interpretation (PARI) method for describing the strategies used by skilled equipment troubleshooters. PARI relies on a structured interview procedure between pairs of experts, commenting on a simulated problem task.

Memory

COGNET (Zachary et al., 1991) is one of the few CTA methods to specifically address the limitations of working memory. COGNET consists of a computer program that is a record of the operator's performance, usually by mapping keystroke analysis into a workload framework; COGNET basically extends the Goals, Operators, Methods, and Selection (GOMS) to include workload limitations as modeled by a blackboard architecture. Few, if any CTA methods address the management of long-term memory. To the extent that mental models suggest memory organization, then the concept mapping and conceptual graph analysis methods would also be relevant here.

RESEARCH ISSUES IN CTA

Improvements in CTA will occur through continued practice and application, but this process will take time. In this section, we describe a range of topics for future research. Some of the topics are linked to the criteria for selecting CTA methods, and other topics are related to the application of CTA methods in general.

How Can We Evaluate CTA Methods?

For a practitioner trying to select a CTA method, there is not a great deal to go on other than its track record of generating discoveries and having an impact. Have boundary conditions been described? Have real-world applications been demonstrated using the method? Are there domains in which the method has been used successfully? Are there domains in which the method has been unsuccessful? Information about validity and utility of various CTA methods is not readily available. There is a critical need for such data, to allow users of various methods to make informed choices. The CTA study must accurately portray the cognitive processes of the specialists being studied. As CTA becomes more visible, demand for evidence of validity and reliability of the methods will and should increase.

How Can We Determine the Validity of a CTA Finding?

Woods (1997), personal communication) has raised the ethical question of how a CTA application will be used. He noted that in many instances, the designers of a system want to eliminate the role of the human, and are seeking to capture the human contribution to the task in a small set of rules that can be turned into an automation of a task. If a CTA application is shallow, and fails to capture the richness and subtlety of the cognitive skills, then it will be easier to dismiss the importance of the human operator, which can lead to serious consequences.

We also need to judge the comprehensiveness of the findings. A CTA project that generates valid conclusions but misses important aspects of cognitive skills, is a failure. When we study the validity of CTA methods, we need to be concerned with what the methods capture and fail to capture, and not just whether the CTA findings are accurate. There have been very few examinations of validity and reliability of CTA methods. Hoffman et al., (1998) describe the systematic attempts to assess validity and reliability of a CTA method, the Critical Decision Method. Militello et al. (1997) developed a paradigm for evaluating the adequacy of knowledge captured by CTA methods. They taught

CTA methods to one group of participants, but not to a second group, and then compared the outputs of the two groups after giving them each a chance to individually interview SMEs. The quality of the outputs (recommendations for cognitive training) were assessed by other SMEs. This type of research is labor intensive, but offers a way to determine the impact of training on the CTA methods. Although these attempts to assess validity of CTA data are encouraging, they do not address the issue of completeness.

How Can We Understand the Boundary Conditions for Specific CTA Methods?

Work here would look at ways to improve the match between the CTA method and the cognitive demands of the task. In the previous section, we offered some initial speculations about the CTA methods that appeared useful for addressing different types of cognitive demands. However, the matching of method to cognitive demands requires that we have a means to distinguish the cognitive demands of a task, and we do not. In the future, we would need a preliminary investigation of a task domain, prior to beginning the knowledge elicitation (see Hoffman, 1987). Following Roth et al. (1992), if we need to specify the cognitive demands of a domain, then that suggests that we will have to first employ some methods for capturing these cognitive demands. Boy (1998) developed a framework for a situational CTA, which he terms a cognitive function analysis. He recommends studying the task, the user, the environment, and the artifact. Vicente (1999) takes a related stance. To accomplish a cognitive work analysis, Vicente recommends that we describe the work domain, the tasks to be performed, the mental strategies needed to perform those tasks, the organizational issues that must be handled, and the operator competencies needed. Anastasi, Hutton, Thordsen, Klein and Serfaty (1997) have made a similar recommendation. They suggested using a method such as an operator function model to specify the cognitive demands, followed by the use of the appropriate CTA method. Roth et al., Boy, Vicente and Anastasi et al. would all seem to argue for an initial CTA step of depicting the task/work domain/environment as a way of understanding the cognitive demands of greatest interest. Then, the knowledge elicitation methods can be selected to match the cognitive demands. We would expect that, as methods are developed for describing the work domain, the tasks to be performed, the functions required, the organizational issues to be handled, and the features of the artifact, then these methods would be used along with CTA methods, in a more comprehensive cognitive engineering study. Work is needed to develop methods that can be used efficiently to describe work domains and environments and can connect these to CTA strategies.

A different perspective on boundary conditions is to evaluate the cost/benefit ratio of CTA methods. Most projects have limited time and money available. Often access to subject matter experts is limited. The potential interviewers or observers may lack experience. These constraints impact the CTA methods used. It will be important to articulate the amount and type of resources required by individual CTA methods. On the one hand, we would like to use a variety of CTA methods to seek converging evidence for our findings. On the other hand, most projects are necessarily limited by time and budget. To transform CTA into a technology, we need to be able to estimate the resource requirements for the methods considered. Hoffman (1987) studied the productivity of different knowledge engineering procedures. Hoffman et al. (1995) reviewed methods used to contrast the efficiency of knowledge elicitation methods. We need additional efforts like these in order to gauge relative efficiency. We appreciate that such studies may be labor intensive. However, if we could establish research paradigms, it might be possible to open the door for graduate student research projects. There are natural tasks, such as driving, playing video games, and using word processing programs, that are within reach of university research efforts and could offer platforms for comparing CTA methods in capturing expertise.

Yet another way of clarifying boundary conditions for CTA methods is by compiling and studying CTA failures to dissect what went wrong. In developing CTA approaches into a technology, one of the barriers is that there is not enough dissemination of failures. Looking at our own experience, we have had our share of CTA projects that were not successful. Sometimes, we failed to achieve the criterion of discovery. Other failures occurred when we were not successful in describing the basis for the expertise. Sometimes, this may have occurred because the so-called experts were not very skilled, as in the example discussed earlier regarding the task of assembling communications equipment. At other times, we probably used the wrong CTA methods. An example would be in trying to identify critical incidents in a situation where the participants didn't have such incidents, or could not remember them. Thus, in a study of baggage screeners (Kaempf, Klinger & Wolf, 1994) we tried to obtain critical incidents, but each screening encounter lasts at most five seconds, so the screeners had no recall of specific events. Fortunately, we were able to shift our methods and rely on observation, with subsequent probing of sample screen images. Ironically, during our observation we were present when a passenger attempted to carry a gun through the screening station, providing an unplanned critical incident. Other failures have occurred because the investigators were not sufficiently skilled. Sometimes this was related to interviewing skills, and sometimes to a lack of background knowledge.

A near failure (see Miller, Militello & Heaton, 1996) occurred when we used CTA methods to examine the way air campaign planners evaluate their plans. The preparation for the interviews identified several different potentially useful CTA strategies. Once the interviews began, one after another of the CTA strategies was applied, and found inadequate. Because of the demonstrated power of asking SMEs to recount real, lived incidents, based on the interviewees' experiences as air campaign planners, Miller et al. (1996) tried to elicit examples of such incidents that they could probe further. As the interviews progressed, the interviewers discovered that air campaign planners are in that position for only a short period of time, three to four years at most. Therefore, they do not have extensive case bases of planning incidents.

In addition, the role of the air campaign planners who were interviewed was to support the Commander in Chief (CINC) in theater. Planning support was delivered to the CINC, who would compile support from several sources. As a result, the air campaign planners received little or no feedback regarding the planning support they had provided. These elements, in addition to the fact that much of their work is highly classified, meant that the air campaign planners had a difficult time recalling and discussing real, lived incidents.

The interviewers had started with the Critical Decision Method, hoping to get the planners to talk about actual incidents. That failed. Then the interviewers used a constructive simulation exercise, asking the planners to design a difficult scenario. That was too unstructured for the planners. Then the interviewers used a probe involving a hypothetical tool that would enable the planners to identify weak links in a plan (which could be examined in greater detail), but that was still too abstract. Next, the interviewers tried a probe whereby the planners had to rely on an assistant to develop a plan; the planner's job was to evaluate the plan. Still too abstract. The fifth and last strategy was to introduce a hypothetical situation that had no plausibility, thereby avoiding concerns of security violations. The planners were asked to imagine that the U.S. had decided to invade Canada, that the planner had developed the plan, and that the plan fell through, because there were too many glitches. What were the glitches? This exercise was concrete enough to engage the planners, yet unrealistic enough to avoid their concerns about classified information.

During this project, in addition to learning about the task of air campaign planning, the CTA team learned a great deal about the boundary conditions for CTA methods that focus on real, lived experience. Had the interviewers gone in with only a single strategy, they would have failed.

How do we Calibrate the CTA Strategy Adopted with the Intended Application or Outcome?

The concern here is that the CTA findings be the appropriate type of representation, at the right level of detail, with the meaningful type of expertise covered. We need to find some way to negotiate with the sponsors and users in advance, so that they receive the product they need. For example, we can save a great deal of disappointment by preparing CTA output samples in advance, and showing these to users in order to determine their preferences. It is important to point out that direct output from knowledge elicitation sessions is rarely a useful means of communicating what has been learned. Analysis, organization, and framing are generally needed to create a meaningful representation of the knowledge captured.

Where are the Guidelines for Human Factors Professionals?

There are several ways to achieve this. We can describe exemplars of good practice. The application of CTA methods requires experience, feedback, and training, just as any type of specialized methodology does. One of the concerns facing the field of CTA is that applications will be disappointing to the sponsors, diminishing enthusiasm. This can occur because too many CTA studies result in massive documentation of irrelevant cognitive details, without making any discoveries, or having any impact. Another reason for loss of enthusiasm will be if too many system designers claimed they are using CTA methods, when they have little training or background. The resultant studies would be likely to be unsuccessful, and the blame could be placed on the methods or the approach, rather than on the analysts. But we don't know how to assess whether someone is qualified or not, or even how long it takes to become qualified on a specific CTA method.

The research questions here include the investigation of individual differences, to see if some practitioners are better or less able to conduct a CTA project than others. Are the same skills needed to use the Critical Decision Method and COGNET and Concept Mapping? Probably not. Therefore, we need to consider individual differences in using specific methods, not differences in CTA per se. For example, in training people to use the Critical Decision Method, we have obtained anecdotal evidence that some very intelligent researchers seem to be insensitive to what counts as a good story, or incident. Others seem to have difficulty in decentering, in taking the perspective of the subject matter expert. If we could demonstrate individual differences, then it would help in selecting personnel to work on CTA projects.

Another research question addresses the types of training that would facilitate competence in applying CTA methods.

Another way to support the wider use of CTA studies would be to derive methods for 'streamlining' the CTA process. Many CTA methods are relatively simple e.g. concept mapping), others can supposedly be learned in just a few hours (conceptual graph analysis) and others have been developed expressly for the purpose of helping untrained personnel get a start at tapping into cognitive skills (Militello & Hutton, 1998). The rationale of these programs is that currently, most designers do not use *any* methods to examine cognitive skills. Therefore, some methods, even simplified ones, are better than none. However, the risk is that people may use these methods carelessly, and then claim that they have captured the essential aspects of expertise needed for a design project, when they have not managed to describe the important cognitive skills. One solution might be to carefully evaluate the validity, reliability, and efficiency of streamlined CTA methods.

CONCLUSIONS

We have defined CTA as the attempt to describe three factors: the cognitive demands presented by the task and the situation, the constraints of the task itself, and the systematic interpretation of the findings. CTA typically attempts to explore aspects of expertise needed for proficient performance, but CTA methods are also used to study the cognitive processes of people who are not proficient, and may be in the process of developing competence.

Currently, a wide variety of CTA methods are being developed and used. As we learn how to define the cognitive demands presented by a task or situation, we hope we will be able to map CTA methods onto these demands, so that we can more efficiently select and apply the appropriate methods. This ability should result in more efficient studies, and greater user satisfaction. It should also help move the field of CTA into becoming more of a technology.

In considering the various steps needed to help develop CTA methods into a technology, one common theme (e.g. Hoffman, 1987) is the need for empirical evaluation of the methods, to determine validity and reliability. We need to understand what the different CTA methods capture, and what they miss. We need to understand the individual differences in analysts that might result in different findings, using the same methods in the same domain. It is not difficult to proliferate new CTA methods. It may be more important now to consolidate, by moving into a period of assessment and review. This type of support may be more available in the research community than in the practitioner community. By helping to assess the utility of different CTA methods, researchers may be able to reach better theoretical descriptions of the nature of expertise, and better descriptions of cognitive skills. The descriptions would have both basic and applied research value.

Finally, some of the greatest impacts from CTA projects occur when discoveries are made about the way complex tasks are performed. This is one of the things that sets CTA efforts apart from conventional task analyses. The exhaustive documentation of precursors and criteria and terminal conditions, the tracing of data pathways for each type of sensor, does not necessarily translate into insights. As we try to turn CTA methods into a reliable technology, we must be careful not to become too fascinated by procedures and objectivity, and lose sight of the importance of discovery. A CTA study is not an extension of hypothesis-testing methods for experimentation. One of the strengths of CTA is to provide a means for generating new hypotheses and unexpected explanations of phenomena, the way people accomplish difficult tasks in natural settings.

ACKNOWLEDGMENTS

We would like to acknowledge the effort of Beth Crandall in reviewing an earlier draft and providing many useful suggestions. We also appreciate the comments and advice of Robert Hoffman and Eduardo Salas.

REFERENCES

Anastasi, D., Hutton, R., Thordsen, M., Klein, G., & Serfaty, D. (1997). Cognitive function modeling for capturing complexity in system design (pp. 221–226). Proceedings of the 1997 IEEE International Conference on Systems, Man, and Cybernetics.

Andriole, S. J., & Adelman, L. (1993). Prospects for cognitive systems engineering, *Proceedings of the 1993 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 743–747). Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Barsalou, I. W. (1985). Ideals, central tendency, and frequency of instantiation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 629–654.

Boy, G. A. (1998). *Cognitive function analysis*. Norwood, NJ: Ablex Publishing Corporation.

Cannon-Bowers, J. A., & Salas, E. (Eds) (1998). *Making decisions under stress: Implications for individuals and team training* (1st ed.). Washington, DC: American Psychological Association.

Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cook, R. I., & Woods, D. D. (1996). Adapting to new technology in the operating room. *Human Factors*, *38*(4), 593–613.

Cook, R., Woods, D., Walters, M., & Cristoffersen, K. (1996). The cognitive systems engineering of automated medical evacuation scheduling and its implication. *Proceedings from the 3rd Annual Symposium on Human Interaction with Complex Systems*. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

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

Cooke, N. J., Durso, F. T., Schwanaveldt, R. W. (1994). Retention of skilled search after nine years. *Human Factors*, *36*(4), 597–605.

Crandall, B., & 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*(1), 42–51.

Crandall, B., Klein, G., Militello, L., & Wolf, S. (1994). *Tools for applied cognitive task analysis* (Technical Report prepared for the Naval Personnel Research and Development Center, San Diego, CA). Fairborn, OH: Klein Associates Inc.

de Groot, A. D. (1946/1978). *Thought and choice in chess* (2nd ed.). New York, NY: Mouton.

de Groot, A. D. (1986). Intuition in chess. *International Computer Chess Association Journal*, *9*, 67–75.

DeMaio, J., Parkinson, S., Leshowitz, B., Crosby, J., & Thorpe, J. A. (1976). *Visual scanning: Comparisons between student and instructor pilots* (Technical Report No. AFHRL-RE-76-10 prepared for Air Force Human Resources Laboratory).

Di Bello, L. (1997). Measuring success in non-trivial ways; how we can know that a DSS implementation has really worked. *Proceedings of the 1993 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 2204–2209). Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Di Bello, L. & Spender, J. C. (1996). Constructive learning: A new approach to deploying technological systems into the workplace. *International Journal of Technology Management*, *11*, 747–758.

Ericsson, K. A. (1996). The acquisition of expert performance: An introduction to some of the issues. In: K. A. Ericsson (Ed.), *The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports, and Games* (pp. 1–50). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Ericsson, K. A., & Smith, J. (1991). *Toward a general theory of expertise: Prospects and limits*. Cambridge: Cambridge University Press.

Essens, P., Fallesen, J., McCann, C., Cannon-Bowers, J., & Dorfel, G. (1995). *COADE = A Framework for Cognitive Analysis, Design, and Evaluation* (Technical Report AC/243 (Panel 8) TR/17). Brussels: Defence Research Group, NATO.

Gopher, D., Weil, M., Bareket, T., & Caspi, S. (1988). Using complex computer games as task simulators in the training of flight skills. *Proceedings of the 1988 IEEE International Conference on Systems, Man, & Cybernetics*.

Gordon, S. E. (1994). *Systematic training programs: Maximizing effectiveness and minimizing liability.* Englewood Cliffs, NJ: Prentice-Hall.

Gordon, S. E., & Gill, R. T. (1997). Cognitive task analysis. In: C. E. Zsambok & G. Klein (Eds), *Naturalistic Decision Making* (pp. 131–140). Mahwah, NJ: Lawrence Erlbaum Associates.

Gordon, S. E., Schmierer, K. A., & Gill, R. T. (1993). Conceptual graph analysis: Knowledge acquisition for instructional system design. *Human Factors*, *35*(3), 459–481.

Gott, S. P., Pokorny, R. A., Kane, R. S., Alley, W. E., & Dibble, E. (1996). Development & evaluation of an intelligent tutoring system: Sherlock 2 – an avionics troubleshooting tutor (Technical Report AL/HR-TR-1996-XX). Brooks AFB, TX.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *Proceedings of IEEE Transactions on Systems, Man, and Cybernetics, SMC-17* (pp. 753–770).

Hanish, K. A., Kramer, A. F., & Hulin, C. L. (1991). Cognitive representations, control, and understanding of complex systems: A field study on components of users' mental models and expert/novice differences. *Special Issue: Cognitive Ergonomics II, Ergonomics*, *34*(8), 1129–1145.

Hayes-Roth, F., Waterman, D. A., & Lenat, D. B. (1983). *Building expert systems*. MA: Addison-Wesley.

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., Crandall, B. W., & Shadbolt, N. R. (1998). Use of the Critical Decision Method to elicit expert knowledge: A case study in cognitive task analysis methodology. *Human Factors*, *40*(2), 254–276.

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* (2), 129–158.

Holding, D. H. (1985). *The psychology of chess skill*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hutchins, E. (1983). Understanding Micronesian navigation. In: D. Gentner & A. L. Stevens (Eds), *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kaempf, G., Klinger, D., & Wolf, S. (1994). *Development of decision-centered interventions for airport security checkpoints* (Technical Report prepared for the U.S. Department of Transportation, Cambridge, MA). Fairborn, OH: Klein Associates Inc.

Kaempf, G. L., Thordsen, M. L., & Klein, G. (1991). *Application of an expertise-centered taxonomy to training decisions* (Technical Report prepared for U.S. Army Research Institute, Alexandria, VA). Fairborn, OH: Klein Associates Inc.

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

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*, 136–153.

Kieras, D. E. (1988) Toward a practical GOMS model methodology for user interface design, In: M. Helander (Ed.), *Handbook for Human-Computer Interaction*. N. Holland: Elsevier Science.

Kirwan, B. & Ainsworth, L. K. (1993). *A guide to task analysis*. London: Taylor and Francis Ltd.

Klein, G. (2000). Cognitive task analysis of teams. In: J. M. C. Schraagen, S. Chipman, & V. Shalin (Eds), *Cognitive Task Analysis* (pp. 417–430). Mahwah, NJ: Lawrence Erlbaum Associates.

Klein, G. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Klein, G. A. (1978). Phenomenological vs. behavioral objectives for skilled performance. *Journal of Phenomenological Psychology*, *9*, 139–156.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fireground. *Proceedings of the 30th Annual Human Factors Society*, *1*, 576–580. Santa Monica, CA: Human Factors & Ergonomics Society.

Klein, G. A., & Crandall, B. W. (1995). The role of mental simulation in naturalistic decision making. In: P. Hancock, J. Flach, J. Caird, & K. Vicente (Eds), *Local Applications of the Ecological Approach to Human-Machine Systems* (Vol. 2, pp. 324–358). Mahwah, NJ: Lawrence Erlbaum Associates.

Klein, G. A., & Hoffman, R. (1993). Seeing the invisible: Perceptual/cognitive aspects of expertise. In: M. Rabinowitz (Ed.), *Cognitive Science Foundations of Instruction* (pp. 203–226). Hillsdale, NJ: Lawrence Erlbaum Associates.

Klein, G. A., Orasanu, J., Calderwood, R., & Zsambok, C. E. (Eds) (1993). *Decision making in action: Models and methods*. Norwood, NJ: Ablex Publishing Corporation.

Klein, H. A., Vincent, E. J., & Isaacson, J. J. (2001) Driving proficiency: The development of decision skills. In: E. Salas & G. Klein (Eds), *Linking Expertise and Naturalistic Decision Making* (pp. 305–322). Mahwah, NJ: Lawrence Erlbaum Associates.

Klinger, D. W. & Militello, L. G. (in preparation). Designing for performance: A cognitive systems engineering and cognitive task analysis approach to the modification of the AWACS weapons director interface.

Layton, C., Smith, P. J., & McCoy, C. E. (1994, March). Design of a cooperative problem-solving system for en-route flight planning: An empirical evaluation. *Human Factors*, *36*(1), 94–119.

Lesgold, A. M., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: diagnosing X-ray pictures. In: M. T. H. Chi, R. Glaser & M. J. Farr (Eds), *The Nature of Expertise* (pp. 311–342). Mahwah, NJ: Lawrence Erlbaum Associates.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review (95)*, *4*, 492–527.

McFarren, M. R. (1987). *Using concept mapping to define problems and identify key kernels during the development of a decision support system.* Master's thesis, School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB, OH.

Means, B., Salas, E., Crandall, B., & Jacobs, O. (1993). Training decision makers for the real world. In: G. A. Klein, J. Orasanu, R. Calderwood & C. E. Zsambok (Eds), *Decision Making in Action: Models and Methods* (pp. 306–326). Norwood, NJ: Ablex Publishing Corporation.

Metcalfe, J., & Shimamura, A. P. (1994). *Metacognition: Knowing about knowing*. Cambridge, MA: MIT Press.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, *104*(1), 3–65.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena. *Psychological Review*, *104*(4), 749–791.

Militello, L. G. (2001). Representing expertise. In: E. Salas & G. Klein (Eds), *Linking Expertise and Naturalistic Decision Making* (pp. 247–264). Mahwah, NJ: Lawrence Erlbaum Associates.

Militello, L. G., & Hutton, R. J. B. (1998). Applied Cognitive Task Analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics, Special Issue: Task Analysis*, *41*(11), 1618–1641.

Militello, L. G., Hutton, R. J. B., Pliske, R. M., Knight, B. J., & Klein, G. (1997). *Applied Cognitive Task Analysis (ACTA) Methodology* (Technical Report prepared for Navy Personnel Research and Development Center). Fairborn, OH: Klein Associates Inc.

Militello, L., & Lim, L. (1995). Patient assessment skills: Assessing early cues of necrotizing enterocolitis. *The Journal of Perinatal & Neonatal Nursing*, *9*(2), 42–52. Gaithersburg, MD: Aspen Publishers, Inc.

Miller, T. E., Militello, L. G., & Heaton, J. K. (1996). Evaluating air campaign plan quality in operational settings. In: A. Tate (Ed.), *Advanced Planning Technology.* Menlo Park, CA: The AAAI Press.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NY: Prentice-Hall.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259.

Pliske, R., Klinger, D., Hutton, R., Crandall, B., Knight, B., & Klein, G. (1997). *Understanding skilled weather forecasting: Implications for training and the design of forecasting tools* (Technical Report No. AL/HR-CR-1997–0003). Brooks AFB, TX: Air Force Materiel Command, Armstrong Laboratory, Human Resources Directorate.

Rasmussen, J., Pejterson, A. M., & Goodstein, L. P. (1994). *Cognitive systems engineering*. New York: John Wiley and Sons, Inc.

Redding, R. E. (1989). Perspectives on cognitive task analysis: The state of the art. *Proceedings of the Human Factors Society 33rd Annual Meeting*, 1348–1352. Santa Monica, CA: Human Factors & Ergonomics Society.

Redding, R. E., & Seamster, T. L. (1994). Cognitive task analysis in air traffic controller and aviation crew training. In: N. Johnston, N. McDonald & R. Fuller (Eds), *Aviation Psychology in Practice*. Brookfield, VT: Ashgate Publishing Company.

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

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits on the search for mental models. *Psychological Bulletin*, *100*(3), 349–363.

Schraagen, J. M. C., Chipman, S., & Shalin, V. (Eds) (2000). *Cognitive task analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Scribner, S. (1985). Knowledge at work. *Anthropology and Education Quarterly*, *16*(3), 199–206.

Seamster, T. L., Redding, R. E., & Kaempf, G. L. (1997). *Applied cognitive task analysis in aviation*. Aldershot, England: Avebury Aviation.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, *53*, 252–266.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129–138.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *American Psychological Association, Inc.*, *119*(1), 3–22.

Smith, E. M., Ford, J. K., & Kozlowski, S. W. J. (1997). Building adaptive expertise: Implications for training design strategies. In: M. A. Quinones & A. Ehrenstein (Eds), *Training for a Rapidly Changing Workplace: Applications of Psychological Research*. Washington, D.C.: APA.

Vicente, K. J. (1999). *Cognitive work analysis: Towards safe, productive, and healthy computer-based work*. Mahwah, NJ: Lawrence Erlbaum Associates.

Woods, D. D. (1993). Process-tracing methods for the study of cognition outside of the experimental psychology laboratory. In: G. A. Klein, J. Orasanu, R. Calderwood & C. E. Zsambok (Eds), *Decision Making in Action: Models and Methods* (pp. 228–251). Norwood, NJ: Ablex Publishing Corporation.

Woods, D. D., O'Brien, J., & Hanes, L. F. (1987). Human Factors challenges in process control: The case of nuclear power plants. In: G. Salvendy (Ed.), *Handbook of Human Factors/Ergonomics*. NY: John Wiley and Sons, Inc.

Zachary, W. W., Zaklad, A. L., Hicinbotham, J. H., Ryder, J. M., Purcell, J. A., & Wherry, R. J. (1991). *COGNET representation of tactical decision-making in ship-based anti-air warfare* (Technical Report 911015.9009 for Naval Ocean Systems Center, San Diego, CA).

Zsambok, C., & Klein, G. (1997). *Naturalistic decision making*. Mahwah, NJ: Lawrence Erlbaum Associates.