Classifying C2 Decision Making Jobs Using Cognitive Task Analyses And Verbal Protocol Analysis

Submitted to the 6th International Command and Control Research Technology Symposium, June 2001, Annapolis, Maryland
Paper Track: C^2 Decision Making and Cognitive Analysis

Thomas R. Gordon, Michael D. Coovert, Dawn L. Riddle, and Donald E. Miles

Institute of Human Performance, Decision Making, and Cybernetics
Department of Psychology, University of South Florida
4202 East Fowler Avenue, BEH 339
Tampa, FL 33620-8200

Kimberly A. Hoffman and Thomas S. King V

Department of Psychology, University of South Florida
4202 East Fowler Avenue, BEH 339
Tampa, FL 33620-8200

Linda R. Elliott

Veridian Engineering
2504 Gillingham Drive, Suite 201
Brooks AFB, TX 78235

Samuel G. Schiflett and Scott Chaiken

Air Force Research Laboratory
AFRL/HEAI, Brooks AFB, TX 78235

Abstract

Weapons Directors (WDs) are cognitively complex, decision making jobs involving air traffic control of friendly assets and tracking hostiles from an airborne platform. Three cognitive task analyses (CTAs) were reviewed for WD jobs on the Airborne Warning and Control System aircraft. 230 tasks were derived from this "meta-CTA." Nine performance categories emerged, representing a cognitive-behavioral model of the WD job domain. To test the model, 38 WDs were observed participating in a computer simulation exercise during which they verbalized aloud their thoughts. These verbal protocols were recorded, transcribed, and reduced to task statements. Using a checklist derived from the meta-CTA, subject matter experts classified each of the statements as either primarily cognitive in nature, behavioral in nature, or a combination of the two. Classifications were analyzed to determine job performance differences between experienced and inexperienced WDs. Verbal protocol analysis presents the opportunity to integrate cognitive task analyses into a job model, yielding a new classification system based on a cognitive-behavior approach. This typology for describing jobs transcends traditional job analysis and has applicability to other complex management jobs involving decision-making, problem-solving, and resource management. More research is needed to validate further the process and identify potential boundary conditions for application.

Introduction

Over the past decade, there has been substantial growth in teams and team work, both in the world of research as well as in industry. This interest is due in part to the increasing complexity of the nature of work. That is, complex tasks lend themselves to using team structures (Sundstrom, De Meuse, & Futrell, 1990). Measuring and modeling team processes is important to understand where and how teams can be employed in the accomplishment of tasks (Dickinson & McIntyre, 1997). It follows that modeling performance and establishing performance criteria in training environments are also important issues (Coovert & Craiger, 1997). Moreover, team competencies (knowledge, skills, and abilities--KSAs) are vital to understanding and evaluating team performance in dynamic environments (Cannon-Bowers & Salas, 1998).

In addition to performing complex tasks, many teams work under stressful, and sometimes-dangerous conditions. These team tasks are usually highly dynamic and require team members to think and act quickly to keep up with rapid environmental changes. One such team is the AWACS/WD team. The AWACS (Airborne Warning and Control System) is an aerial platform that serves as an airborne command post. It contains a number of suites with specialized missions in the areas of command, control, and communications (C^3). Among these suites are the Weapons Directors (WDs), who control three aspects of the air battle: (1) high value airborne assets such as refueling tankers and the AWACS itself, (2) fighter aircraft which provide air cover and protection against enemy aircraft, sometimes referred to as “CAP” – combat air patrol – or “DCA” – defensive counter-air -- and (3) fighter-bombers whose mission is to attack surface targets posing a threat to the overall mission of the command; the latter are called “strike” aircraft. The dynamic nature of the environment, the complexity of team tasks, and the criticality of flawless performance combine to suggest the WD tasks ideal for the employment of teams. In this introduction, we describe previous work in order to set the stage for reviewing initial and subsequent phases of a research effort that specifically addresses WD teams.

A number of team performance functions are important to effective team coordination (e.g., information exchanges among team members, monitoring, backup, etc.) (Fleishman & Zaccaro, 1992). Better anticipation of each other’s actions and reduced, more efficient communication among team members can also result in better team performance (Klimoski & Mohammed, 1994). Cannon-Bowers, Salas, and Converse (1993) suggested that mental models encoded (and shared) by team members are particularly important to team performance.

Identifying the WD Performance Domain

A key change in decision theory has involved the increased use of naturalistic decision theory, whereby highly trained observers study complex human decision-making processes in naturalistic settings. Additionally, researchers in human-computer interaction (HCI) and other domains within cognitive science have been developing and utilizing a variety of cognitive task analyses (CTAs), in order to "analyze and model the cognitive processes that give rise to human task performance in specific domains" (Zachary, Ryder, &

Hicinbothom, 1998, p. 315). Thus, the purpose of this paper is show how CTA can be used to describe jobs in a way that goes beyond traditional job analysis.

Seamster, Redding, Cannon, Ryder, and Purcell (1993) describe a cognitive task analysis of en route air traffic controllers, whose functions are similar to those of WDs. For their study, CTA involved identifying key tasks and analyzing those tasks "from a cognitive perspective, identifying the task sub-goals and the triggers that activate the individual task" (p. 260). The authors placed special emphasis on the relation of the tasks to the expert's mental model, while focusing on the planning processes and maintenance of situational awareness.

Miles et al. (2000) conducted research to analyze the functional team and the specific positions of WDs working within the AWACS aircraft. The goal of this research was to develop enough knowledge and understanding of the WDs' positions (and environment) to use as the basis for a model that would integrate both cognitive and behavioral task analyses. Although cognitive and behavioral analyses have been modeled separately, an integrative model, which links these two aspects, has eluded researchers. The difficulty in developing this type of model lies in the limitations of conventional methodologies. Specifically, task concurrencies and dynamic behavior pose a barrier to conventional tools, which can only handle static and synchronous properties. The problem with conventional tools has been exacerbated by cognitive task analyses/behavioral task analyses generally focusing on the individual rather than the team level. Therefore, Miles, et al. conducted a qualitative "meta-analysis" using data gathered from previous cognitive task analyses in order to develop a working model of team performance within the dynamic AWACS environment. This effort resulted in a list of 230 tasks contained in the working model. Through a series of sorting exercises, the 230 WD tasks were reduced to a set of nine working categories. Below is a list of the nine categories and brief descriptions.

Category	Description
A Priori Knowledge.	This category specifies the types of knowledge that WDs must bring to each mission. Much of this knowledge comes from experience and training, but it also includes mission-specific knowledge gathered from the pre-mission briefing.
Individual Internal Cognitions.	This involves the broad types of cognitive activities that AWACS WDs engage in on the job. This category differs from some of the others (e.g., pre-mission planning and resource management) in that it is not usually tied to a specific function or goal. Rather, it addresses the requisite cognitive activities within and across many different aspects of job performance.
Situational Awareness.	This category involves maintaining an understanding of current and projected events involving aircraft in the Area of Responsibility (especially cognizance of the "Big Picture") in order to determine (or predict) events that will require WD activities.
Resource Management.	We view this category as a goal, which WDs strive to achieve. Resource Management involves utilizing AWACS resources to an appropriate extent (i.e., maintaining enough resources while avoiding waste and overload).

Pre-Mission Planning.	This category involves the various tasks that occur during the pre-mission planning process.
Directing.	This is viewed as the primary WD responsibility. The WD directing tasks are pre-prioritized according to Rules of Engagement, orders from higher authority, and Standing Operational Procedures – usually under the premise that the more critical the event, the greater the cognitive load on the WD. We sub-divided the directing category according to the extent to which the directing tasks impact operator cognitive load.
Dynamic Operational Planning and Prioritizing.	This category includes the planning and prioritizing tasks that occur "on the fly," during the actual mission. This dimension involves reassessment of the situation, problem-solving, and reclassification.
Team Issues.	This category involves tasks and events that require or are otherwise related to teamwork, including member communication, coordination of tasks, and cooperation.
Communication.	This final category refers to the act of transmitting or exchanging information which occurs: during pre-mission planning; as part of dynamic operational planning and prioritizing; during directing; expressly for team issues; and to maintain situational awareness. It includes the specific protocols for communication.

The nine categories are not exhaustive, nor are they mutually exclusive. Additionally, they are not hierarchical in nature. They were developed as a useful aid in creating a model to examine performance of WDs within an AWACS team. Once an accurate framework for performance was defined, then better specified models could be developed.

WD Model Development

The follow-on initiative was outlined by Gordon et al. (2000), in which a model of AWACS team performance was proposed and demonstrated. Models of team performance are generally consistent in considering three levels of analysis: individual performance, team performance, and organizational performance and support (Brannick & Prince, 1997). In this respect, team performance is considered in the traditional systems model, using input, throughput, and output as ways to judge important aspects of teams' operations, successes, and failures (Bertalanffy, 1968).

Gordon, et al. (2000) examined team performance in the context of environmental complexity, using cognitive task analytic methodology to link cognitive and behavioral components of team performance (van der Schaff, 1993), then modeling this performance dynamically – that is, in a way that captured the complexity of team performance. Researchers were able to simulate linking explicitly the cognitive categories to operator behaviors, using Petri nets. To date, access to a simulation or platform that allowed linking cognitions and behaviors temporally has not been available. With the future development of more event-based scenarios, we should be able to demonstrate rather easily the efficacy of the Petri net as a modeling tool for this purpose. Any scenario can be analyzed by applying the performance categories to aspects of trigger events and examining their interactions.

Verbal Protocol Analysis

Described first in considerable detail in a classic book by Ericsson & Simon (1984, 1993), verbal protocol analysis (VPA) uses participants' own verbal reports as data. It is used principally to explore cognitive processes. The authors made explicit the techniques to be used and – most important – placed the technique on firm ground by addressing the nature and reliability of the method. In its simplest form, verbal protocol involves recording orally communicated "thoughts," committing them to written form and comparing them with observed behavior.

The overarching objective of the studies described (Miles, et al., Gordon, et al., 2000) was to offer a method toward better understanding of team performance in a dynamic environment through CTA and modeling. These efforts were successful within the limitations posed by available tools and allowed for further study of the WD team. These studies also provided the foundation for development and testing of the hypotheses outlined in the next section. The purpose of the latest phase of the WD study was to combine CTA and VPA to develop a better, more complete picture of WD performance.

Research Design

Participants

Participants consisted of Weapons Directors (WDs) undergoing simulation training at an Air Force Base, in the southwest U.S. There were 17 "experienced" and 21 "inexperienced" WDs, for a total of 38. All had at least minimum qualifications achieved from basic WD training. There were four females and 34 males. The distinction between an experienced and inexperienced WD is important. WDs were categorized as experienced if they had more than 400 flight hours in the AWACS aircraft and at least one year designated as "combat mission ready." All others were categorized as inexperienced WDs. These distinctions are based on established policies of the 552nd Training Squadron for WDs assigned to WD positions. Because of limitations with the platform and scenario, only six of the nine performance categories were considered (pre-mission planning, communication, and team issues were not included).

Procedure

Each participant underwent a four-hour session on the simulation. The first two hours were spent receiving instruction on the simulation and practicing. After a break, participants performed two high intensity scenarios (parallel forms) on which they received individual scores. Following these two graded sessions on the scenario platform, participants were trained on the "think aloud method" (van Someren et al., 1994). Participants were then presented with one 30-minute high intensity computer simulation during which audio recordings were made for later analysis. Transcriptions were made and subsequently reduced to a series of single statements by a subject matter expert who previously performed WD functions in the Navy. A checklist was compiled from Air Force documents to use in rating these statements (referred to as the Task Rating List), and was affirmed by a panel of subject matter experts previously involved in the meta-CTA described earlier.

The Task Rating List was used by a group of volunteer graduate and undergraduate students at the University of South Florida who were trained in its use. The purpose was to rate each of the statements made by the WDs and classify them as primarily behavioral, cognitive, or a blend of cognitive and behavioral. Subsequent to rating the classes, they were asked to choose from among the tasks in the Task Rating List the three-digit code most nearly describing the activity or cognition. The three digit code consists of the performance category (first digit), sub-category (second digit), and the specific task (third digit). Pairs of raters were asked to rate the statements independently and then meet to compare and achieve consensus. All consensus ratings were subsequently reviewed independently by two subject matter experts before they were ultimately recorded. Thus each statement was rated four times both as to selecting the class (primarily cognitive, primarily behavioral, or a blend of the two) and the three digit task code.

The purpose of the ratings was to affirm or disconfirm the hypotheses that predicted, primarily on the basis of experience level, how WDs were expected to express their behaviors and cognitions as they performed their duties in a simulated task environment. In general, we expected experienced WDs to talk more about their activities as they proceeded through the scenario because they had performed these activities more often than the inexperienced WDs in the past.

Analysis

Hypotheses were tested as indicated below.

Hypothesis 1: More experienced WDs will obtain higher individual scores than inexperienced WDs, as measured by the scoring algorithm contained in the computer simulation platform.

The scenario platform provides total scores on individual, team, and mission tasks for each participant. Because the simulation was executed using only individual players, only individual scores were collected. Mean scores were calculated for experienced and inexperienced WDs on all three simulations on which scores were available and compared using a t-test for independent samples.

Hypothesis 2: In verbal protocol analysis, experienced WDs will make more statements (from which cognitions can be inferred) in each of the six performance categories than will less experienced WDs, as measured by the frequency of hits on the task rating list.

Hypothesis 3: Likewise, experienced WDs will make more statements (from which behaviors can be inferred) in each of the six performance categories than will less experienced WDs, as measured by the frequency of hits on the task rating list.

Hypothesis 4: Moreover, experienced WDs will make more statements (from which a blend of cognitions and behaviors can be inferred) in each of the six performance categories than will less experienced WDs, as measured by the frequency of hits on the task rating list.

Hypothesis 5: There will be statistically significant differences between experienced and inexperienced WDs, where total inferences of cognitions, behaviors, and a blend of the two, based on the total number of statements derived from verbal protocol analysis collapsed across the six cognitive categories.

The frequency data available to analyze hypotheses 2 through 5 are nominal data. The appropriate method in this instance is the chi-square (χ^2) test of independence, generally used to test the relationship between two discrete variables. This procedure was used to test hypotheses 2 through 5.

The Sample

Of the 38 WDs from whom data were gathered, a review of the verbal protocols showed that only 31 were usable. Those deemed unusable were primarily due to equipment malfunction, although four WDs simply could not “get the hang of it” in terms of thinking aloud.

Results

Scoring Performance

The platform on which the scenario was run has the capability of calculating scores for mission accomplishment, scores for team performance, and individual scores based on predetermined algorithms built into the simulation. For this particular study, the Air Force determined that WDs would play only as individuals (specifically as the Defensive Counter-air WD, or DCA). Thus the only performance scores displayed were for individuals.

Hypothesis 1

Mean individual scores, associated sample sizes, and standard deviations are shown in Table 1. These scores reflect all trials on the simulation (a total of three), not just the single trial where think aloud verbal protocol) was employed. Note there were little apparent differences between experienced and inexperienced WD mean scores.

Table 1
Individual performance scores as a function of experience level

Experienced			Inexperienced			All Participants		
n	M	SD	n	M	SD	n	M	SD
21	140.67	6.68	17	138.99	5.78	38	139.92	6.28

An independent samples t-test was conducted; no significant differences were revealed between experienced and inexperienced WDs on individual performance scores ($t_{(36)}$=0.828, n.s.). This finding is identical to that of Hoffman (2000) who worked with essentially the same data in a somewhat different context. Accordingly, hypothesis 1 was not supported.

One of the interesting issues uncovered during this study was the lack of objective differences between those WDs classified as experienced vs. those classified as inexperienced. For example, in our study population, on average inexperienced WDs have 224 flight hours in the AWACS, with over 13 months of experience as a WD. This is compared with some 1,400 flight hours for the experienced WDs on average and over 7 years experience. While the differences are statistically significant, the training, simulation hours, check rides and the like probably blur the distinctions, particularly on a relatively simple simulation as this one. The mean scores were clearly not different between the experience levels; moreover, the maximum score that could be achieved on the scenario in use was 152 points. It is highly likely that a ceiling effect was at work. We must look beyond individual scores to find meaningful differences between experienced and inexperienced WDs.

Hypothesis 2

Table 2 is the contingency table for a chi-square test of independence for statements classified as primarily cognitive in nature. Note there are a total of 3,467 statements so classified from among the 31 WDs considered in this analysis.

Table 2
Chi-square contingency table for *cognitive* statements as a function of experience level

Weapons Directors	No. of Statements (Observed Cell Value)		Expected Cell Value	χ^2
Inexperienced (n=16)	(51.61%)	1,768	1,789	
Experienced (n=15)	(48.39%)	1,699	1,677	
Column Totals	(100.00%)	3,467	3,467	.52

The critical value of χ^2 is 3.84, df=1, $\underline{p}<.05$. There is no statistically significant difference between the inexperienced and experienced WDs with regard to the frequency of their statements classified as principally cognitive in nature; i.e., experienced WDs do not make more such statements than do inexperienced WDs. Accordingly, hypothesis 2 is not supported.

Hypotheses 3, 4, and 5

Table 3
Chi-square contingency table for *behavioral* statements as a function of experience level

Weapons Directors	No. of Statements (Observed Cell Value)		Expected Cell Value	χ^2
Inexperienced (n=16)	(51.61%)	558	616	
Experienced (n=15)	(48.39%)	635	577	
Column Totals	(100.00%)	1,193	1,193	11.18

The critical value of χ^2 is 6.64, df=1, $\underline{p}<.01$. The hypothesis is supported (i.e., experienced WDs make more statements of a behavioral nature than do inexperienced WDs).

Table 4
Chi-square contingency table for a *blend* of cognitive and behavioral statements as a function of experience level

Weapons Directors	No. of Statements (Observed Cell Value)		Expected Cell Value	χ^2
Inexperienced (n=16)	(51.61%)	525	539	
Experienced (n=15)	(48.39%)	520	506	
Column Totals	(100.00%)	1,045	1,045	.79

The critical value of χ^2 is 3.84, df=1, $\underline{p}<.05$. The hypothesis is not supported (i.e., there is no difference between experienced and inexperienced WDs).

Table 5
Chi-square contingency table for *total* statements as a function of experience level

Weapons Directors	No. of Statements (Observed Cell Value)		Expected Cell Value	χ^2
Inexperienced (n=16)	(51.61%)	2,851	2,944	
Experienced (n=15)	(48.39%)	2,854	2,761	
Column Totals	(100.00%)	5,705	5,705	6.12

The critical value of χ^2 is 3.84, df=1, $p<.05$. The hypothesis is supported (i.e., experienced WDs make more total statements than do inexperienced WDs).

The ratio of cognitive statements to behavioral statements for experienced WDs was 2.68; for inexperienced WDs the ratio was 3.17. A comparison of these ratios suggests that experienced WDs perform less cognitive work than their inexperienced colleagues.
Further investigation was undertaken to determine if there were meaningful differences in the task codes used by experienced and inexperienced WDs within each class. Results are shown in tables below.

Cognitive Statements

Table 6
Top five cognitive statements by task code for *experienced* WDs

Number of Statements	Code	Definition	Example
175	32C	Situational Awareness: Monitor assets available for mission accomplishment	Out of my 15 alpha, I've got almost 6,000 pounds of fuel – fighters are due at 0530.
127	41G	Resource Management: Monitor resources available.	Check their states before I send them out.
112	63A	Individual Internal Cognitions: Interpret events in the environment.	Looks like it's doing something, maybe it's launching Tomahawk cruise missiles.
106	32B	Situational Awareness: Make judgments about what to do with aircraft.	Let's see the B's – the B's we're gonna drive all the way to the forward end of the island, 'cause they have enough fuel.
97	62A	Individual Internal Cognitions: Perceive events in the environment.	There are six targets coming out of the north.

Table 7
Top seven cognitive statements by task code for inexperienced WDs

Number of Statements	Code	Definition	Example
132	62C	Individual Internal Cognitions: Attend to relevant information.	O.K. J-Stars moved south. There's now 4 hostile tracks in the north.
125	41D	Resource Management: Track the headings of both friendly and hostile aircraft.	Tanker's still on his way northbound and I've got less than a minute before my next set of fighters comes up.
121	41G	Resource Management: Monitor resources available.	Checkin' fuel states on all my guys now at this time, makin' sure I'm good to go.
114	63A	Individual Internal Cognitions: Interpret events in the environment.	(That) looks like a couple of bombers were destroyed – lookin' good.
105	32C	Situational Awareness: Monitor assets available for mission accomplishment.	O.K. Killed the westernmost bomber, so F15B-makin' sure he's committed on the current westernmost bomber.
104	64A	Individual Internal Cognitions: Evaluate formulated plan (or plans).	Everyone is committed at this time, the fight looks good.
100	32B	Situational Awareness: Make judgments about what to do with aircraft.	Let's see – there's another A-1 aircraft I can employ.

A comparison of tables 6 and 7 confirms that, experienced WDs talk about the same as inexperienced WDs as they go through the simulation (on average 113 cognitive statements for experienced WDs vs. 110 cognitive statements for inexperienced WDs). They differ considerably, however, in their focus of attention. Experienced WDs attend first to maintaining situational awareness and do so substantially more than their less experienced colleagues (281 statements vs. 205). Inexperienced WDs tend to talk more about individual internal cognitions (350 vs. 209), suggesting that they are less apt to react quickly and automatically, probably due to a lack of experience with the situations they are facing in the simulation. From a qualitative perspective, experienced WDs focus on *monitoring*; inexperienced WDs focus more on *tracking*. It also appears that experienced WDs do a more efficient job of information fusion.

Behavioral Statements

Table 8
Top six behavioral statements by task code for *experienced* WDs

Number of Statements	Code	Definition	Example
128	22C	Directing: Commit a friendly aircraft to a hostile aircraft (high demand, routine event).	Not good. Let's take 15B and commit him to that Mig-21.
79	21J	Directing: Coordinate fighters to man CAP (combat air patrol) points (routine event).	We got foxtrot up north with a weapon load out of five slammers each, so I'm having him set a CAP.
56	21I	Directing: Direct friendly assets (routine events).	15B – we're sending him up to the north right now.
46	22B	Directing: Direct fighters to intercept with tankers (high demand routine event).	I'm sending Delta to get some gas.
42	21E	Directing: Vector friendly aircraft to any point in 3D space (routine event).	All right, send him over there to this point.
39	41I	Resource Management: Access information on an aircraft.	And it's time to check the gas on F15D, F15D down to 14,000 pounds – he's still very good.

Table 9
Top five behavioral statements by task code for inexperienced WD

Number of Statements	Code	Definition	Example
151	22C	Directing: Commit a friendly aircraft to a hostile aircraft (high demand, routine event).	I just had two fighters heading out to the west – engage the six group….
62	21J	Directing: Coordinate fighters to man CAP (combat air patrol) points (routine event).	Set Charlie up forward CAP.
55	21I	Directing: Direct friendly assets (routine events).	Moving my C's to the north here.
42	22B	Directing: Direct fighters to intercept with tankers (high demand routine event).	F-15C is low on gas, so I'm committing him to the tanker.
36	41I	Resource Management: Access information on an aircraft.	F-15D – checking his status –see what kind of armaments he has.

A comparison of these tables shows that experienced WDs talk slightly more (on average 42 behavioral statements for experienced WDs vs. 35 behavioral statements for inexperienced WDs). The focus of attention for both is on directing and managing resources, major components of the WD task. Qualitatively, it is apparent that inexperienced WDs require more time and expend more effort on directing assets. In addition, from a qualitative standpoint, experienced WDs appear more efficient at information fusion. There may also be a degree of automaticity in responses on the part of more experienced WDs.

Blended Statements

Table 10
Top four blended statements (combination of behavior and cognition) by task code for *experienced* WDs

Number of Statements	Code	Definition	Example
43	32B	Situational Awareness: Make judgments about what to do with an aircraft.	Since there are no (hostile) fighters at this point, I'll go ahead and commit my fighters on the bombers that are up.
38	22C	Directing: Commit a friendly aircraft to a hostile aircraft (high demand, routine event).	F-15B out in the east, now committed on the last couple of bombers, and I will also commit him on some more bombers coming up.
35	41G	Resource Management: Monitor resources available	Fighter Bravo has 18,000 pounds still, fighter Alpha 5,150.
35	41I	Resource Management: Access information on an aircraft.	F-15C's gotta be getting low on fuel – will check his fuel status – he's got about 5,300 pounds.

Table 11
Top five blended statements (combination of behavior and cognition) by task code for *inexperienced* WDs

Number of Statements	Code	Definition	Example
55	32B	Situational Awareness: Make judgments about what to do with aircraft.	Split them up because I've really right now only got F-15B on the west – correction, I mean on the northeast.
45	22C	Directing: Commit a friendly aircraft to a hostile aircraft (high demand routine event).	And we've got a Mig-23, so I'm gonna immediately commit F-15D on the Mig-23.
41	41G	Resource Management: Monitor resources available.	Checking Alpha's fuel – it should be fine.
29	32C	Situational Awareness: Monitor assets available for mission accomplishment.	(Referring to aircraft he lost track of)- There they are, they returned to base.
27	21J	Directing: Coordinate fighters to man CAP points (routine event).	And now the Alpha fighter to a more southern CAP to start off.

WDs talk about the same (on average, 35 blended statements for experienced WDs vs. 33 blended statements for inexperienced WDs). The focus of attention for both categories was on maintaining situational awareness, directing, and managing resources. In terms of frequencies, clearly the inexperienced WDs spend more time and effort on situational awareness. Qualitatively, it takes more effort for inexperienced WDs to accomplish the same tasks that are accomplished in the same timeframe by their more experienced colleagues. Again, this argues for more efficient processing of information and the likelihood of a degree of automaticity of responses by experienced WDs.

Discussion

Summary of Results

The original goal of this study was to develop a methodology or process appropriate to model performance based on a meta-cognitive task analysis and to show how cognitions and behaviors are linked in highly complex environments. The process was developed by using the Air Force Weapons Director job as an exemplar of the kind of position which is highly complex, contains elements of behaviors and cognitions, and would lend itself to examination as to whether the process had merit. A great deal was achieved and there is much to build on from what has been learned in the study.

Findings

Beginning with a meta-cognitive task analysis (Miles, et al.; Gordon, et al., 2000), which resulted in a model containing nine performance categories, a major thrust of the study plan was to examine differences between experienced and inexperienced weapons directors (WDs) on as many of those nine categories as feasible. It was hypothesized that experienced WDs would score higher than inexperienced WDs on a computer simulation exercise. This hypothesis was not supported; i.e., there were no differences between the groups in terms of their performance on the simulation. A partial explanation for these results probably lies in the degree of difficulty of the scenario. There was a substantial ceiling effect on the scores, with almost all WDs, regardless of their level of experience, achieving near maximum scores.

One area in which differences between experienced and inexperienced WDs were found was in the verbal protocol analysis. Originally intended to serve as a means to develop cognitions and to show links with behaviors observed in the simulation, the method proved highly informative for examining predicted differences between the experienced and inexperienced WDs. It was hypothesized that experienced WDs would render more statements through verbal protocol than would their less experienced colleagues in several classifications. Specifically, it was predicted that statements dealing principally with cognitions, and behaviors, and a combination or blend of the two, would be significantly greater in frequency for experienced WDs. These predictions were only partially supported.

Perhaps the most important finding, however, is how the two groups differ on another aspect of the verbal protocol analysis. WDs tend to talk *differently* as a function of their experience level. With respect to statements principally of a cognitive nature (Tables 6 and 7), experienced WDs expend considerably more energy in maintaining situational awareness than do their less experienced peers; the inexperienced WDs appear to devote more energy to internal cognitions (getting their bearings, remembering to look around and see what needs doing). Qualitatively, the experienced WDs appear to be more relaxed and comfortable with the simulation and the tasks to which they must attend than do the inexperienced WDs.

The differences are not quite so striking with the more behaviorally-oriented statements. Both groups spend the bulk of their energy directing friendly assets and managing their resources, both functions that are taught early to new WDs in training. These are the "bread and butter" issues that all WDs must do to be successful, so it should come as no surprise that much time and effort are devoted to them.

The blend classification proved very difficult for the task raters; it may be that there was something in the manner in which the WDs expressed themselves where a clear distinction was not made between the cognitive and behavioral that made their statements more difficult to interpret, that is, to place in the cognitive or behavior classification rather than resorting to the blend classification. Investigating this would make an interesting study for someone so inclined, trained not only in verbal protocol analysis, but also in speech analysis.

Potential Limitations

This study was conducted in a military context, with resources not always readily available and affordable in the civilian world. Nevertheless, scaled back versions of the process could easily be adapted and could reap great dividends for those willing to take a minimal risk. Any complex job that has both cognitive and behavioral components can be subjected to the process.

Small sample sizes do not necessarily adversely impact a study when one is doing what essentially amounts to knowledge engineering, that is, mining knowledge from experts. Using small samples is fairly common.

True subject matter experts (e.g., WDs) did not make the judgments concerning how to classify the statements and what task designators to assign. Instead, we used university students. This is potentially a legitimate criticism. However, the students employed in this endeavor were highly motivated, well trained, and unbiased in their judgments.

There was only one scenario, necessitating an attempt to assess all of the performance categories without sufficient events to do so. As noted earlier, this is a legitimate concern. Continued efforts should be made to expand the opportunities to test all aspects of the model as a way to validate the process.

Theoretical and Practical Implications

The application of verbal protocol analysis in this study was key to what has been learned. There was no other practical way to get at what the WDs were thinking. Using verbal protocol analysis is strongly recommended for future studies of this type. However, a word of caution is in order.

Verbal protocol analysis is manpower intensive and very time-consuming. In the present study, fourteen people worked on the analysis in some fashion. Taping the participant, transcribing the tape, breaking down the transcriptions into single statements, analyzing them with four different judges (classifying the statements and assigning a task from the task rating list), and entering data into the computer takes an enormous amount of time and effort. For each of the WDs, it is estimated that 18 hours were spent on each tape from start to finish, meaning that some 558 hours were spent just on the 31 tapes. An effort of this magnitude should be carefully designed and executed to avoid wasting valuable resources, especially people's time. The research team discovered some automated programs advertised to conduct protocol analysis; however, these programs basically grouped like terms without considering the context that human judges could consider.

The USF AWACS Performance Model

Although not a major study objective, great strides have been made toward validating the WD performance model, especially in the areas of Directing, Maintaining Situational Awareness, Resource Management, and Internal Cognitions. The simulation and its scenario used in the present study precluded the incorporation of Pre-Mission Planning, Communication, and Teamwork. Dynamic Operational Planning and A Priori knowledge were also not included due to the nature of the simulation.

In the future, the AWACS Study Team anticipates the opportunity to use different platforms and simulations. When those are event-based, it is reasonable to assume that further progress can be made in the validation of the performance model, due in large measure to a wider range of potential responses to scenario events. In the meanwhile, much has been learned about what WDs do. Moreover, enough has been learned that modifications to the model can be made with reasonable certainty that they will reflect empirically grounded, theoretically sound explanations of how WDs do their work.

Application Beyond the WD Context

It appears that the methodology developed in the present study will have considerable utility in other sectors. Demonstrating how to establish a model that is both descriptive and prescriptive through cognitive task analysis and verbal protocol analysis, and then testing it, is a major contribution of this study. It would appear useful in a broad array of organizations well beyond the WD context.

For example, a manager charged with the responsibility for doing strategic planning in which he or she must undertake to integrate a number of cognitively complex tasks would likely find this procedure useful. The model could be used in selecting new team members based on measuring varying amounts of expertise on the elements defined by a model similar to that developed for WDs. People could be trained to the standards specified by the model; it could also be used as a diagnostic tool and for providing specific types of feedback. In this regard, developing a model could provide the means for establishing meaningful criteria for performance appraisal.

Knowing the intricacies and complexities of a series of complex tasks which go to make up many jobs could assist in another way. In coming years it is reasonable to anticipate that a number of these complex tasks will lend themselves to performance, at least partially, by intelligent agents. Testing has already begun on providing an agent to assist in the performance of a WD's job (Hoffman, 2000). Knowing the cognitive complexities of the job, beyond that which could be provided by traditional job analysis, would greatly advance the development of these intelligent agents.

Some Conclusions and Final Thoughts

This study has provided a new classification system not previously discussed in the literature. Verbal protocol analysis has given us the opportunity to integrate cognitive task analyses into job models – in effect, we have a new typology for describing jobs that goes a step beyond the more traditional KSAs and other task-oriented tools. Job analysts have generally concentrated on behaviors associated with the job, while knowledge engineers basically ignored behaviors in favor of cognitions. With the present study, there is a good chance that members of the community can be persuaded to adopt the notion of looking at cognitive task analyses, integrating them with verbal protocols, and examining jobs in the context of the classifications described in this study (cognitions, behaviors, blends). These classifications are, of course, defined on the basis of a series of elements relevant to the job, such as our task rating list provides. A very important issue, which is fodder for future research, involves looking at trade-offs among cognitions, behaviors and blends. Put another way, we need to improve the operational definitions describing what is contained in a cognitive, behavioral, and blended statement, and examine more fully the amount of overlap among them. It is also strongly suggested that more research be undertaken to validate the process we have begun here and to recommend additional strategies or approaches that would improve it. A good start in this regard is provided in a recent work by Vicente (1999) in which cognitive work analysis is described. It appears that methodologies from our efforts can easily be integrated with many of the methodologies described by Vicente.

References and Bibliography

Barr, A., & Feigenbaum, E.A. (1982). The handbook of artificial intelligence. Los Altos, CA: William Kaufmann, Inc.

Bertalanffy, L. (1968). General systems theory: Foundations, development, applications. New York: George Braziller.

Brannick, M.T. & Prince, C. (1997). An overview of team performance. In Brannick, M.T., Salas, E., & Prince, C. (Eds.). Team performance assessment and measurement: Theory, methods and applications. Mahwah, NJ: Lawrence Erlbaum Associates.

Buchanan, B.G., & Feigenbaum, E.A. (1978). DENDERAL and Meta-DENDERAL: Their applications dimension. Journal of Artificial Intelligence, 11, 5-24.

Canon-Bowers, J.A., & Salas, E. (1990). Cognitive psychology and team training: Shared mental models in complex systems. Symposium presented at the 5th Annual Conference of the Society for Industrial and Organizational Psychology, Miami, FL.

Cannon-Bowers, J.A., Salas, E., & Converse, S.A. (1993). Shared mental models in expert team decision making. In N.J. Castellan, Jr. (Ed.), Current issues in individual and group decision making (pp. 221-246), Hillsdale, NJ: Lawrence Erlbaum Associates.

Cannon-Bowers, J.A., & Salas, E. (1998). Teamwork competencies: The interaction of team member knowledge, skills, and attitudes. In H.F. O’Neil (Ed.), Workforce readiness: Competence and assessment. Mahwah, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121-152.

Cobb, M.G. (1998). The impact of environmental complexity and ambiguity on AWACS crew performance in simulated combat scenarios. The Pennsylvania State University Graduate School: Unpublished Master’s Thesis.

Coovert, M.D., & Craiger, J.P. (1997). Modeling performance and establishing performance criteria in training systems. In Ford, J.K. (Ed.). Improving training effectiveness in work organizations. Mawah, NJ: Lawrence Erlbaum Associates.

Coovert, M.D., Craiger, J.P., & Cannon-Bowers, J.A. (1995). Innovations in modeling and simulating team performance: Implications for decision making. In Guzzo, R.A., Salas, E. (Eds.). Team effectiveness and decision making in organizations. San Francisco: Jossey-Bass.

Coovert, M.D., & Dorsey, D.W. (2000). Computational modeling with Petri nets: Solutions for individual and team systems. In Hulin, C. & Ilgen, D. (Eds.). The third discipline of scientific research in psychology: Computational modeling and computer simulation of behavior. San Francisco: Jossey-Bass.

Dalrymple, M.A. (1991). Evaluating Airborne Warning and Control System strategy and tactics as they relate to simulated mission events. Technical Report AL-TP-1991-0049. Armstrong Laboratory, Brooks AFB, TX.

Dickenson, T.L., & McIntyre, R.M. (1997). A conceptual framework for teamwork measurement. In Brannick, M.T., Salas, E., & Prince, C. (Eds.), Team performance and measurement: Theory, methods, and applications. Mahwah, NJ: Lawrence Erlbaum Associates.

Elliott, L.R., Cardenas, R., & Schliflett, S. (1999, June). Development and analysis of team-based rating scales of AWACS performance in distributed mission training. Proceedings of the Annual Research and Technology Symposium on Command and Control, Newport, RI.

Elliott, L.R., Schliflett, S.G., Hollenbeck, J.R., & Dalrymple, M. (In press). Situational awareness and team performance in realistic command and control scenarios. In M. McNeese, E. Salas, & M. Endsley (Eds.), Group situational awareness: New views of complex systems.

E-3 AWACS in Service Worldwide (1999). [Online]. Available http://www.com/defense-space/infoelect/e3awacs.index.htm

Ericsson, K.A., & Simon, H.A. (1984, 1993). Protocol analysis: Verbal reports as data. Cambridge, MA: MIT Press.

Elliott, L.R., Schifflet, S. G., & Dalrymple, M.A. (1998). Modeling the decision making process and performance of airborne warning and control systems (AWACS) weapons directors. Paper presented at the 1998 NATO Symposium on Collaborative Crew Performance in Complex Operational Systems, Edinburgh, United Kingdom.

Fahey, R.P., Rowe, A., Dunlap, K., & DeBoom, D. (2000). Synthetic task design (1): Preliminary cognitive task analysis of AWACS weapons director teams. Technical Report AFRL-HE-AZ-TR-2000-159. Brooks AFB, TX: Armstrong Laboratory.

Fleishman, E.A., & Zaccaro, S.J. (1992). Toward a taxonomic classification of team performance functions: Initial considerations, subsequent evaluations and current formulations. In R.W. Sweezy and E. Salas (Eds.), Teams: Their training and performance (pp. 31-56). Norwood, NJ: Ablex.

Gordon, T.R., Riddle, D., Hoffman, K.A., Miles, D., King. T.S., Coovert, M.D., Foster, L.L., Elliott, L.R., & Schiflett, S. (2000). Team performance: Integrative dynamic models linking cognitive and behavioral components. Poster session presented at the annual meeting of the Society of Industrial and Organizational Psychology, New Orleans, LA.

Hoffman, K.A. (2000). Trust and performance with intelligent agent technology: Implications for human-agent interaction. Unpublished manuscript, University of South Florida.

Kleinman, D.L., & Serfaty, D. (1989, April). Team performance assessment in distributed decision making. Paper presented at the Simulation and Training Research Symposium on Interactive Networked Simulation for Training, University of Central Florida, Orlando.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? Journal of Management, 20 (2), 403-437.

Klinger, D.W., Andriole, S.J., Militello, L.G., Adelman, L., & Klein, G. (1993). Designing for performance: A cognitive systems engineering approach to modifying an AWACS human computer interface. Air Force Technical Report AL/CF-TR-1993-0093, Air Force Armstrong Laboratory, Crew Systems Directorate, Human Engineering Division: Wright-Patterson AFB, OH.

Levine, J.M., & Moreland, R.L. (1990). Progress in small group research. Annual Review of Psychology, 41, 585-634.

MacMillan, J., Serfarty, D., Young, P., Klinger, D., Thordsen, M., Cohen, M., & Freeman, J. (1998). A system to enhance team decision-making performance (AP-R-1102). Woburn, MA: Aptima, Inc.

Miles, D.E., Hoffman, K.A., Foster, L.L., King, T.S., V, Gordon, T.R., Riddle, D., Coovert, M.D., Elliott, L.R., & Schiflett, S. (1999). A qualitative methodology for integrating cognitive task analysis data. Unpublished manuscript, University of South Florida at Tampa.

Minionis, D.P., Zaccaro, S.J., & Perez, R. (1995). Shared mental models, team coordination, and team performance. Paper presented at the 10th annual meeting of the Society for Industrial and Organizational Psychology, Orlando, FL.

Reisig, W. (1992). A primer in Petri net design. Berlin: Springer-Verlag.

Rouse, W.B., & Morris, N.M. (1986). Understanding and enhancing user acceptance of computer technonoly. IEEE Transactions on Systems, Man, and Cybernetics, 16(6), 965-973.

Rumelhart, D.E., & Ortony, A. (1997). The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro, & W.E. Montague (Eds), Schooling and the acquisition of knowledge (pp. 99-136). Hillsdale, NJ: Lawrence Erlbaum Associates.

Seamster, T.L., Redding, R.E., Cannon, J.R., Ryder, J.M., & Purcell, J.A. (1993). Cognitive task analysis of expertise in air traffic control. International Journal of Aviation Psychology, 3, 257-283.

Shiflett, S., Strome, D., Eddy, D., & Dalrymple, M. (1990). Aircrew evaluation sustained operations performance (AESOP): A triservice facility for technology transition (USAFSAM-TP-90-26). Brooks AFB, TX: USAF School of Aerospace Medicine.

Shiflett, S., & Elliott, L.R. (2000). Synthetic team training environments: Application to command and control aircrews. In H. F. O'Neil, Jr., & D. Andrews (Eds.). Aircrew Training and Assessment.

Sundstrom, E., DeMeuse, K.P., & Futrell, D. (1990). Work teams: Applications and effectiveness. American Psychologist, 45, 120-123.

Swezey, R.W., & Salas, E. (Eds.). (1992). Teams: Their training and performance. Norwood, NJ: Ablex.

van Someren, M.W., Barnard, Y.F., & Sandberg, J.A.C. (1994). The think aloud method. San Diego, CA: Academic Press, Inc.

van der Schaaf, T.W. (1993). Developing and using cognitive task typologies. Ergonomics, 36(11), 1439-1444.

Veldhuyzen, W., & Stassen, H.G. (1977). The internal model concept: An application to modeling human control of large ships. Human Factors, 19, 367-380.

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

Zachary, W.W., Ryder, J.M., & Hicinbothom, J.H. (1998). Cognitive task analysis and modeling of decision making in complex environments. In Cannon-Bowers, J. & Salas, E. (Eds.). Making decisions under stress: Implications for individual and team training. Washington, DC: American Psychological Association.

Zurada, J.M., Marks, R.J., & Robinson, C.J. (Eds.) (1994). Computational intelligence imitating life. New York: IEEE Press.