ARI Research Note 98-29

Measuring Battlefield Knowledge Systems: Test of a Protocol Analysis Approach

Rex R. Michel
U.S. Army Research Institute

Ft. Leavenworth Research Unit
Stanley M. Halpin, Chief

August 1998

U.S. Army Research Institute
for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the U.S. Total Army Personnel Command

EDGAR M. JOHNSON
Director

Technical review by

Sharon L. Riedel

NOTICES

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED (from. . . to)
August 1998	Final	June 1996 – June 1998

4. TITLE AND SUBTITLE Measuring Battlefield Knowledge Structures: Test of a Protocol Analysis Approach	5a. CONTRACT OR GRANT NUMBER 5b. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Rex R. Michel (ARI)	5c. PROJECT NUMBER 633007A 5d. TASK NUMBER H01 5e. WORK UNIT NUMBER 1141
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: TAPC-ARI-RK Fort Leavenworth Research Unit Building 90, McClellan Ave Fort Leavenworth, KS 66027-1360	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. MONITOR ACRONYM ARI 11. MONITOR REPORT NUMBER Research Note 98-29

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 words)*:

This document reports on the design and testing of a method and measures for evaluating the battlefield domain knowledge structures of US Army officers. The method and measures were based on a review of the literature on expertise summarized in this report. Three different tactical situations were presented to 31 Army Officers with varying levels of relevant tactical experience. Each situation had a different response requirement intended to provide various measures of the integrity of the officer's knowledge base and the levels of abstraction and application it contains. The response verbal protocols were analyzed to derive the measure values. The results indicated a generally negative relationship between experience and high scores on the knowledge structure measures contrary to expectations. Possible reasons for the findings are discussed.

15. SUBJECT TERMS

Battlefield expertise Knowledge structures Protocol analysis Performance measurement

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	27	Rex Michel DSN 552-9790

MEASURING BATTLEFIELD KNOWLEDGE STRUCTURES: TEST OF A PROTOCOL ANALYSIS APPROACH

CONTENTS

LIST OF TABLES

LIST OF FIGURES

MEASURING BATTLEFIELD KNOWLEDGE STRUCTURES: TEST OF A PROTOCOL ANALYSIS APPROACH

Introduction

Need and Objectives

The age of automation has ushered in a near flood of innovative training methods. For the United States Army, many of these innovations have involved the use of simulations to train both small and large units in realistic battlefield situations.

The effectiveness of simulation-based training is typically measured by observation of external behaviors and structured questioning of trainees and trainers (e.g., Shlechter, Shadrick, Bessemer & Anthony, 1997). But many of the most robust and critical effects on individuals are cognitive in nature and not readily assessed by such methods. Nor are they adequately measured by the typical classroom question-and-answer examination. The reason for this is that the primary cognitive effect of learning by experience is an increased understanding of the relationships among objects, events and actions given particular situations. What is needed is a means of assessing the gain in this "operational understanding" as a result of simulation training.

Such a measurement instrument would also be a useful research tool to test various learning interventions and in cognitive studies of expertise and individual differences. It might also provide an effectiveness measure for individual-based battlefield simulation training techniques; a growing interest in the United States Army.

Therefore, the objectives of this study were to:

(a) Design a method for measuring the knowledge structuring effects of experience-based learning, drawing from the literature on domain expertise and related cognitive subjects.
(b) Test the measurement method using Army officers with a wide range of experience to determine if it discriminates among levels of experience.

Overview of Literature

Over 130 articles and book chapters relevant to expertise and knowledge-related issues were reviewed and pertinent information summarized or extracted over the course of this research and related projects. The summarized references were again reviewed to eliminate information that did not directly relate to the problem at hand. This resulted in consideration of 60 of the original references. This, by no means, represents an exhaustive search of the literature; in the past two decades hundreds of empirical and theoretical works have been produced on these topics. In fact, most of the works cited here were ones published six or more years ago. It is felt, however, that they are representative of major themes and findings in the field important to our research project.

The relevant information from these 60 references was categorized into 12 topic areas having some relevance to the description, use, and measurement of experience-based knowledge. Of those 12, five topic areas are briefly summarized here, the rest either being subsumed under them or proving to be of less importance to the project than originally thought.

Conceptual differences. Attempting to measure the structure of domain knowledge rather than its sheer amount assumes that changes occur in the way that knowledge is put together in the mind as practical experience in the domain increases. There is ample evidence in the research literature to support this assumption.

Most researchers agree that not only does the amount of information grow with experience in a domain, but also the knowledge becomes better organized (Glaser, 1984; Ceci & Ruiz, 1992; Royer, Cisero, & Carlo, 1993; Federico, 1995). Organizational changes in the knowledge base are in at least two general directions. One type of change is the gradual addition of more and more abstract knowledge. Many studies comparing expert to novice performance found that novices tend to view situations in terms of the concrete objects presented or in terms of simple, concrete procedures or details. Experts tend to view situations in terms of abstract principles or general domain concepts (Wiser & Carey, 1983; Scribner, 1986; Lawrence, 1988; Johnson-Laird, 1989; VanLehn, 1989). An example of research reflecting this difference found that altering the game of bridge differentially affected experts and novices—changes in the deep-structure rules of the game affected experts more than novices while surface changes affected the novices more (Sternberg & Frensch, 1992).

The second general type of knowledge structure change with experience is that the knowledge base becomes more interrelated. Many different theoretical systems have been proposed for explaining and representing this linking of knowledge in human memory (Rumelhart & Norman, 1985). A basic assumption of most of these systems, whether they be concepts, chunks, productions, schemas, scripts, frames, networks, or mental models is that through experiences an individual associates events and objects together so he or she is able to respond quickly and appropriately to complex sets of stimuli without having to consciously think through every step in the process (Schank, 1985; Simon, 1985; Johnson-Laird, 1989; Rumelhart, 1989; Smith, 1989; Anderson, 1990). Much of what relates otherwise isolated pockets of facts together is knowledge of how to apply that information in various situations—knowledge of how actions are related to specific situations (Norman, 1985; Rumelhart & Norman, 1985; Lesgold, 1988). Rumelhart (1989) suggests comparing the number of 'natural type' entities mentioned in a subject's protocol with the number or 'role type' (action-related) entities mentioned—this is analogous to measuring the percentage of entities that have an action element in them. The assumption is that the ability to apply knowledge reflects how well the knowledge is integrated.

The combination of more abstract knowledge and greater connectivity of knowledge means the expert is able to make more inferences than the novice. In fact, in highly technical fields, it is impossible for a naïve individual to make a strong, principled commitment to a particular interpretation (diSessa, 1983). The expert, on the other hand, has the abstract knowledge and knowledge connectivity to detect complex but familiar patterns in the presented data (Clancy, 1988) or to infer properties and features of a problem not present in the problem presentation (Groen & Patel, 1988; Reimann & Chi, 1989).

General problem solving strategies. When novel problems are faced within a domain, novices and experts tend to use similar general problem solving strategies (Glaser, 1984, Reimann & Chi, 1989, VanLehn 1989). However, the expert's superior knowledge base and organization even in relatively novel situations permits generation of more or better hypotheses and more thorough tests of those generated (Glaser, 1984; Voss & Post, 1988; Reimann & Chi, 1989; Foley & Hart, 1992; Federico, 1995). Expert/novices differences are more apparent in the performance of routine or common tasks in the domain. Here the expert works efficiently and with little apparent effort applying domain-specific, knowledge-based, content-dependent strategies to rapidly arrive at the solution (Soloway, Adelson & Ehrlich, 1988; Humphreys, et al, 1990; Royer, Cirero & Carlo, 1993). It follows that the problem solving strategies of experts are apt to differ between "hard" and "easy" problems, thus eliciting different knowledge structures, especially in reference to procedural knowledge (Foley & Hart, 1992).

Association/recall as a primary problem solving strategy. Expert performance on "easy" problems is generally attributed to memory. Fischoff (1988) makes the point that people in general interpret what they see whenever it is even remotely possible--stimuli from the environment activate associations with a wide network of related events stored in memory. For a domain expert, this association/recall process can retrieve a large network of relevant memories, both conceptual and procedural, often without any conscious awareness of the process (Miller, 1985; Glaser & Chi, 1988; Groen & Patel, 1988; Lesgold, et al, 1988; Posner, 1988; Staszewski, 1988; Ericsson & Simon, 1993).

Several specific studies within the military domain support this phenomenon. In researching rapid recognitional decision making, Klein, for example, has found that it is most apt to happen when the military decision maker is experienced, time pressure is greater, and conditions are unstable (Klein, et al, 1990). Solick, et al (1997) found that the influence of experience on accuracy of predicting battle outcomes was dependent on the inherent predictability of the scenario. Experienced officers did better on a normal mission plan; they were less accurate on one that was poorly executed. Similar findings have also occurred in other domains (Foley & Hart, 1992).

Recent innovative methods of studying brain functions and chemistry have greatly expanded our understanding of how these associations are made (Fischbach, 1992; Goldman-Radic, 1992; Damasio, 1994). This burgeoning field of research may eventually answer many of the questions that remain about skilled memory and recall.

Domain task-specific measurement. There is a considerable amount of research suggesting that the concept of a "general" intelligence may be a myth (Sternberg, 1990; Gardner, 1993; Ceci, 1996). Glaser (1984) even suggests that skilled performance on aptitude and intelligence tests is the result of the exercise of conceptual and procedural knowledge in the context of specific knowledge domains. His research suggests that learning and reasoning skills are not abstract mechanisms, but the result of the transfer of this "conditionalized knowledge" to other domains.

In fact, any single person has a wide range of cognitive strengths and weaknesses and high skill areas are more apt to be born from long periods of study and practice than from general aptitude (Staszewski, 1988). Therefore, to assess an individual's skill, we need to measure it in the domain of interest under conditions where they are performing tasks normally required in the domain (Royer, Cisero & Carlo, 1993).

Research has found that the use of task simulations reduces the correlation between practical and academic intelligence to almost zero (Wagner, 1986). There is further research suggesting that the superior knowledge organization of the expert cannot be measured in the absence of actual domain task performance (Bellezza, 1992). Also, people do not make the types of inferences that distinguish expertise in the absence of a triggering mechanism—the presentation of some event or relationship in the domain that cues previously unrelated knowledge sets (Holyoak & Nisbett, 1988). For example, Voss & Post (1988) found that when they gave experienced political scientists a hypothetical problem in their domain, they used their knowledge to construct plausible causal factors, even to the point of constructing a plausible "history" for the event. This reinforces the point made by Royer, Cisero & Carlo (1993) that if you only measure the amount of declarative knowledge people have mastered, it does not give an indication of where they are along the skill development continuum. They may be novices who have memorized a list of the steps, but whose actual performance is still slow and error-prone.

Protocol analysis. For tasks with a high cognitive element, one of the few options available for recording task performance is to ask the subject what they are thinking while performing the task. A primary issue in obtaining these verbal protocols is how much the data collector should interact with the subject during the elicitation. Ericsson & Simon (1993) indicate that requiring subjects to explain their utterances is likely to altar their cognitive process. They also state that verbal protocols should be obtained concurrent with task performance as only then will the subject be responding to the thoughts that are driving performance in response to specific cues.

The purpose of the research, the data required, and the difficulty of obtaining rich spontaneous protocols, however, often make it essential that the researcher interact with the subject. For example, subjects are often asked what their objectives are or why they reached particular conclusions (Clancy, 1988; Lawrence, 1988; Bibby, 1992; Forsythe & Barber, 1992). As verbal protocols typically tap but a very small portion of a subject's domain knowledge, experimenters sometimes ask questions to see what other knowledge the subject possesses (Kuipers & Kassirer, 1984). Clancy (1988) asked why subjects did not ask certain questions in order to determine if assumptions were being made.

How to categorize responses is another issue in protocol analysis. One of the biggest problems is simply defining the boundaries between "separate" thoughts

(Johnson, 1988; Fletcher & Huff, 1990). This can be a nontrivial task because even though a complex thought may be activated as a unit from long-term memory, the requirement to verbalize it makes it appear as a sequence of propositions due to the limited capacity of short-term memory (Ericsson & Simon, 1993).

Categorizing the types of verbalizations is a highly individualistic process usually driven by the intent of the research and the theoretical bent of the researcher (Johnson, 1988; Fletcher & Huff, 1990). However, a useful distinction between declarative and procedural knowledge (i.e., objects and relationships) is frequently made (Kuipers & Kassirer, 1984; Groen & Patel, 1988; Duff, 1992; Forsythe & Barber, 1992). Concept mapping is often used to diagram this basic distinction. Another frequently made distinction is between utterances which merely parrot back the original problem stimuli and those that reflect some cognitive processing (Groen & Patel, 1988). Utterances reflecting cognitive processing may be further broken down into those that 'paraphrase' the stimuli and those that reflect inferential reasoning or a chain of inferences (Frederiksen, 1986; Ericsson & Simon, 1993). Linking of related utterances is typically done by grouping them into "chains of inference" or "arguments" that are some times diagrammed as IF-THEN production statements (Groen & Patel, 1988; Lawrence, 1988; Fletcher & Huff, 1990).

The types of measures taken from verbal protocols typically involve frequency counts for categories like those mentioned above (Frederiksen, 1986; Groen & Patell, 1988; Forsythe & Barber, 1992). Time to respond and length of utterances are also sometimes recorded (Ericsson & Simon, 1993). Other measures that require some judgement on the part of the researcher are fairly common in protocol analysis. Examples include checking for errors and inappropriate ways of dealing with stimuli (Lawrence, 1988; Duff, 1992), looking for indicators of knowledge that was not verbalized (Kuipers & Kassirer, 1984), distinguishing between forward and backward reasoning (Groen & Patel, 1988), and differentiating between deep causal structure reasoning and specific situation schematic reasoning (Bibby, 1992). Another measure sometimes used is the location and length of hesitations in speech, whether they be pauses, repetitions, or nonsense utterances such as "ahhh" (Rochester, et al, 1977; Ericsson & Simon, 1993). It is hypothesized that hesitations are indications of shifts in processing of cognitive structures. One interpretation is that hesitations within major cognitive themes indicates a word-choice or lexical decision; between major themes it represents decisions concerning the overall direction of thought or syntactic structure.

Hypotheses

Based on findings from the literature review and the objectives of this research, the following hypotheses were derived.

- H1: As amount of task-related experience increases, the integrity of one's knowledge bases as measured by the number and quality of identified relational propositions also increases. This is the primary knowledge structure measure. It is assumed that it reflects the strength of association among concepts in the mind.

- H2: There is no significant relationship between amount of task-related experience and the number of attributes that can be identified. As stated in the review of the conceptual differences literature, the primary effect of experience on knowledge is not amount, but organization of what is known. Several studies have found that journeymen are just as good at recalling specific facts (and sometimes better) as experts (for example, see Groen & Patel, 1988). If this is the case, attributional knowledge, or specific facts about objects, should not vary significantly among our test participants as a result of experience.
- H3: As amount of task-related experience increases, the level of abstraction of identified characteristics increases as measured by the proportion of implicit to explicit characteristics identified. This hypothesis acknowledges the hierarchic nature and generalizability of expert knowledge structures as reviewed in the literature.
- H4: Performance on directed-response measures of knowledge structures is related to performance on the same measures in a nondirected tactical problem. If our measures are valid, then they must equate to performance on tasks that are more realistic than our constrained "laboratory" tasks.

There is a basic assumption here that the criterion measure used in this research of amount of relevant job experience is at least an adequate approximation of level of expertise. There are many factors that contribute to the development of expertise, but it was assumed that combat arms Army officers who hold tactical positions in tactical units are building more integrated knowledge bases over time. This experience factor alone should be strong enough to produce significant relationships with knowledge structure measures.

Method

The methods used to elicit, record and analyze knowledge structures were highly dependent on information gained from the literature review. Therefore, the following principles guided the design.

a.) Use domain-specific task samples as basis of measurement.
b.) Use concurrent verbal protocols to elicit cognitive measures.
c.) Measure the degree of abstraction of responses.
d.) Measure the degree of interrelationship of the verbal protocol.

Instrument and elicitation

Three different sets of domain-specific stimuli were developed, two were battalion level situations and the other a brigade situation. Each one consisted of a single graphic display of a tactical situation with minimal or no verbal description. Each situation contained only three of the METT-T factors (Mission, Enemy, Terrain, [own] Troops, and Time) typically used by the military to analyze and describe a tactical situation. For example, one situation (ETT) displayed only the enemy, terrain, and own troops with no mention of the mission. A second one (MTT) displayed the mission, terrain, and own troops but not the enemy. The third (MET) contained the mission, enemy, and own troops but no terrain. None of the three contained any timing information. This was done in an attempt to force the participant to make assumptions about the missing METT-T factors. If successful, these assumptions should help determine the level of abstraction at which the participant is reasoning as well as how closely his knowledge sets are interrelated.

Each participant was given all three tactical situations. For each tactical situation the participant was given a different response requirement. The combinations of tactical situations and response requirements were counterbalanced across participants. The order of tactical situations was varied, but the order of response requirements remained constant. For the first tactical situation presented, the participant was allowed to study the situation as long he wanted. Then the graphic display was taken away and he was asked to describe the tactical situation in his own words from memory. For the second tactical situation presented, the participant was asked to describe the important attributes for each METT factor shown in the graphic. For the final tactical situation presented, the participant was asked to describe the important relationships between each pair of METT factors present and then to describe any relationships that took into account all three of the displayed METT factors. For the second and third presentations, the participant was allowed to retain the graphic while responding. It was felt that the use of free memory response and attributional and relational directed responses would provide the range of responses needed to sample a participant's knowledge structure.

When all three problems were completed, the three graphic displays were laid out in front of the participant and he was asked which one was the easiest to respond to, and then which one was the hardest to respond to. Participants were also asked why it was the easiest or hardest. This was done to provide a subjective measure of the interaction of scenario and treatment effects as an estimation of the cognitive load of each. If the type of scenario had an effect on cognitive load, it would moderate the desired treatment effect.

After administering the three limited scope problems, each participant was given a more complete tactical problem. This was a fictitious Desert Storm problem that had been used previously in other projects. Much more information was given and the participant was asked to develop a concept for how he would respond to this more complex situation. This scenario was to be used to see how well knowledge structure scores on the simpler problems predicted similar scores on a more complex, realistic, problem.

In addition to the tactical problems, each participant filled out a background questionnaire. This questionnaire asked for their rank, time in service, time in grade, military schooling and the title, echelon and length of all tactical command and staff positions they held throughout their career. This information provided the independent variable, amount of relevant experience, which is hypothesized to predict the degree of structuring within an officer's domain knowledge base.

Measures of knowledge structure

Simple counts were taken of the number of entities mentioned for each tactical situation/response requirement pair. For the free response from memory ("Describe") condition, this was the number of situation characteristics mentioned regardless of type. For the Attribute condition it was the number of attributes. For the Relationship condition it was the number of relationships and the number of characteristics per relationship (see Figure 1).

A count was made of the number of errors that occurred in reporting entities. Because of the subjectivity inherent in defining errors, the only ones recorded were perceptual errors. These were errors in Level 1 entities as defined in Table 1 below.

The data reduction performed on the verbal protocols retained the order in which entities were verbalized. This order was available for analysis based on the assumption that knowledge entities most readily associated with the stimuli will be reported first. This provides another insight into the structure of domain knowledge.

The level of abstraction of individual entities in a protocol was operationally defined using three distinct categories. The category definitions are contained in Table 1. Counts and proportions for each level were recorded.

Table 1

Operational Definitions of Levels of Abstraction

Level	Title	Operational Definition
1	Perceptual Response	Counts Relative and absolute locations Repeating stimulus words without adding meaning Naming a displayed object Simple comparisons ('our A to their B') without value judgment
2	Direct Inference	Direct attribute of a presented object (other than name) Naming missing objects or information
3	Indirect Inference	Requires a chain of inferencing that may or may not be verbalized Attributes of an object not presented

The number of relationships identified in the relational response condition was not deemed a sufficient indicator of the important measure of knowledge base integrity. To provide a measure of the integrity of the participant's knowledge base across all three response requirements, the linking of entities within each protocol was recorded. This included the number of links, the percent of entities linked, the average and greatest link depth and the average number of entities per link. Within the relational response condition, the quality of each identified relationship was judged as either high or low and these percentages were also recorded.

From the background questionnaires four separate measures of experience were taken. These included time in service, total months in tactical positions, total months in tactical units, and total months at echelons equal to or greater than battalion level. The last measure was required because the three tactical situations were at battalion or brigade level.

Protocol reduction

All verbal protocols were audio recorded and transcribed into written form. Excel worksheets were created for separating the individual entities in the protocols and to indicate the nesting (i.e., linking) of entities. Figure 1 is given as an example of the worksheets. The figure is a facsimile of the first sheet of the Relational condition for the first six subjects in the experiment.

Knowledge Structures Analysis Worksheet—Relationships (TacSit#3)

	ETT	MTT	MET	ETT	MTT	MET
Sequence	FH01	FH02	FH03	FH04	FH05	FH06
1	EN/TERRAIN	MSN/TERRAIN	MSN/OWN TRPS	EN/TERRAIN	MSN/TERRAIN	MSN/OWN TRPS
0.1	Obviously coming down ave of appch	We're on W side of the river	No task organiztn being conducted	Chosen to use hgh spd ave of appch .	Secure the river is extremely diffcult	Should have enough cbt pwr to do msn
0.01	from standpoint of S2	must secure rvr from atk from E	two task forces	running E and W	there is hgher grd on other side of rvr	msn is des en in zone
0.001	doing an IPB		atking in zone	thru Cntrl Crridor	en atking frm there	en is an MRB plus
0.0001					must secure frm W	
0.0002					or move far forwrd	
0.00001					to secure on E side	
0.002	and their COA		forward			you have bde cbt tr
0.0001						gives 3 to 1 cbt pwr relationship
0.00001						should be enough
0.000001						assuming all other factors are equal
0.003	and ave of app					
0.0001	and branches					
0.02			armor battalion			
0.001			following			
0.002			in reserve			
0.03			operating pure			
0.001			which is strange			
0.2	Atleast 1 sct sect staying off main ave	As far as terrain goes	Prbably need more artillery	Stuck to N Wall		
0.01	as should do	wooded areas	an arty bn	to max masking		
0.001		either side of a major road	supporting	from what he thinks is the sgnfcnt threat		
0.0001		comng thru cnter of our sector	makes sense	our AT co		
0.0002			but this being the force			

Figure 1. Facsimile of matrix worksheet.

For the Relational as well as Attributional conditions, the major categories discussed were given by the experimenter (e.g., "What are the important relationships between the enemy and the terrain in this situation?"). The second level in the Relational condition (.1, .2, etc. shown in Figure 1) represents the relationships identified by the participant. Levels below that represent linked/nested detail given by the participant. As with the other two response conditions, the number of entities per link and the depth level of each link were calculated.

In Figure 1, each numbered comment represents a separate entity as determined by the experimenter. Each entity for each response condition was annotated as to its appraised level of abstraction, whether or not it represented an action, and whether it contained a perceptual error.

Participants

Participants came from three Army posts within the continental United States. All were Army officers assigned to the experiment during scheduled research weeks at their posts. Adequate recordings were obtained for 31 of the 32 officers participating. The relevant demographics for these 31 officers are contained in Table 2. All were males.

Table 2

Demographics of Participants

Rank		Branch		Time In Service (Months)	
1st Lieutenant	3	Combat Arms	29	Range	28 - 228
Captain	7	Combat Support	2	Mean	150.75
Major	18			Standard Dev	53.6
Lieutenant Colonel	3				
Time In Tactical Positions (Months)		Time In Tactical Units (Months)		Time In Units => Bn (Months)	
Range	0 – 141	Range	0 – 141	Range	0 - 92
Mean	56.37	Mean	60.72	Mean	30.25
Standard Dev	34.45	Standard Dev	35.95	Standard Dev	25.87

Results

The participants' subjective judgments as to which scenario was easiest and which hardest to respond to produced significant results. Table 3 shows that the response requirement was a significant determinant of how participants judged the scenarios.

Table 3

Participant Judgments of Scenario Response Difficulty

By Scenario

	Easiest	Neither	Hardest
MET	11	5	11
MTT	5	13	9
ETT	10	11	6

Chi Square = 7.432 (p =.115)

By Response Requirement

	Easiest	Neither	Hardest
Describe	14	9	4
Attributes	9	12	6
Relations	3	8	16

Chi Square = 17.435 (p =.002)

From the table it appears that the Describe condition was the easiest, the Relations condition the hardest and Attributes somewhere in between. It is interesting to note that only three of the 27 participants who judged difficulty named the response requirement as a reason for their judgment. Yet the analysis indicates a high probability that it was an important factor. Most participants mentioned the missing METT factor as the important determinant with different participants stating the same missing factor alternately as the cause for a 'easy' or 'hard' judgment.

Table 3 indicates no discernable significant effect of the particular scenario on judged difficulty of responding. However, the scenario used did have a significant effect on many of the knowledge structure measures. It turns out that the MET scenario, which contains no terrain information, was far less productive than the other two scenarios on several measures. Of 16 measures used, the MET scenario had a significant depressive effect on five of them ($p<.05$ on either a difference of variance or difference of means test). The MET scenario showed a "tendency" toward depressing productivity on five other measures ($p<.15$). As might be expected, all four of the quantity measures used were among those affected.

Pearson r correlation was the primary statistic used to analyze the results. As might be expected, there were generally highly significant relationships among the individual measures comprising a particular response requirement (of 18 intra-treatment paired measure correlations, Pearson r's with $p<.005$ were obtained on 16 of them). This high interdependency suggests that one or two of these measures per response requirement would adequately test whatever the set is measuring. In fact, about 46% of the inter-treatment paired measure correlations were significant at $p<.05$. This indicates a fairly low degree of independence among the measures in general.

The surprise came in looking at the relationship between the knowledge structure measures and the criterion measures of experience. Only nine of 48 paired measure correlations here (12 knowledge structure x 4 experience measures) were significant and these were all negative. In fact, almost all the correlations (43 of 48) were negative. Controlling for the scenario effect did increase the number of significant correlations ($p<.05$) to 17, but all of them were negative (all four quantity measures remained nonsignificant). These results indicate that as an officer grows in experience, he tends to use less abstraction, less depth, and less action in his description of a tactical situation; at least as measured by this research.

Table 4 shows the Pearson r correlations between the six measures unaffected by the scenario effect and the four experience measures. These data are indicative of the general results.

Table 4

Correlation of Selected Knowledge Structure Measures With Experience Measures

	Knowledge Structure Measures					
	Describe		Attributes		Relations	
Experience Measures	% Abstract Level 3	% Action	Average Depth	% Abstract Level 3	Average Depth	% Abstract Level 3
Time In Service	-.24	-.23	-.38*	-.45**	-.20	-.03
Time in Tac Pos	-.10	-.08	-.37*	-.22	-.04	-.17
Time In Tac Unit	.08	.11	-.43**	-.18	-.20	-.20
Time In ≥ Battalion	-.02	-.05	-.47**	-.23	-.24	-.24

*p<.05 **p<.01

Note the generally negative correlations in Table 4. An analysis of selected scattergrams of the distributions does not indicate any nonlinear relationships that might explain this. It should be noted that there is a fair degree of skewness ($>\pm.5$) in four of the six knowledge structure measures and two of the four experience measures.

The knowledge structure measure with all four significant negative correlations with experience is the Attributes response requirement measure of average depth of attribute descriptions. This measure is also the most affected by experience of all 16 knowledge structure measures. Participants with the most experience tended to spend less time flushing out or justifying the attributes of the situation that they identified.

Table 5 lays out the scores for all 31 participants on the six measures that were not affected by the scenario effect.

Table 5

Participant Standardized Scores Across Six Knowledge Structure Measures In Ascending Order of Mean Score Plus Associated Experience Scores

Knowledge Structure Measures										
Describe		Attributes		Relations			Experience Measures			
% Ab Lvl 3	% Action	Avg Depth	% Ab Lvl 3	Avg Depth	% Ab Lvl 3	Mean Score	Mos In Service	Mos In Tac Pos.	Mos In Tac Unit	Mos In Unit≥Bn
.67	.77	1.00	1.00	1.00	.90	.890	28	2	2	2
.58	.36	.88	.81	.76	1.00	.732	66	32	32	0
.97	.89	.89	.71	.80	.10	.723	156	84	102	40
.47	.30	.66	.67	.43	.82	.558	153	60	80	44
1.00	.45	.66	.66	.07	.43	.545	72	29	34	13
.00	.00	.97	.88	.63	.72	.533	180	18	18	11
.38	.14	.67	.60	.49	.84	.520	120	48	60	30
.43	.07	.64	.45	.77	.61	.494	204	92	56	39
.52	.35	.44	.58	.34	.72	.492	186	77	96	42
.00	.00	.75	.82	.54	.77	.480	48	34	34	1
.93	1.00	.00	.32	.13	.26	.439	156	72	106	34
.33	.36	.49	.41	.28	.69	.427	180	24	36	36
.32	.07	.40	.51	.48	.64	.403	228	90	112	80
.23	.09	.48	.47	.26	.75	.380	108	33	45	15
.22	.13	.84	.74	.00	.34	.378	180	32	28	6
.00	.00	.49	.23	.51	.90	.355	168	36	36	4
.00	.00	.63	.36	.34	.67	.333	192	72	72	12
.00	.07	.32	.21	.67	.67	.323	180	123	75	21
.00	.00	.31	.38	.46	.74	.315	216	111	90	51
.28	.20	.24	.30	.26	.56	.307	155	66	99	71
.00	.20	.33	.29	.44	.56	.303	165	83	83	65
.21	.11	.23	.19	.15	.77	.277	204	60	105	46
.13	.14	.18	.00	.29	.87	.268	204	0	0	0
.00	.00	.42	.51	.16	.33	.237	48	36	36	8
.08	.00	.21	.23	.46	.41	.232	108	0	4	4
.43	.00	.30	.14	.14	.34	.225	168	52	38	38
.00	.00	.47	.21	.22	.44	.223	186	72	72	53
.05	.00	.26	.18	.31	.51	.218	156	50	50	20
.00	.00	.04	.30	.15	.56	.175	189	141	141	92
.13	.00	.32	.12	.19	.10	.143	84	32	32	13
.21	.00	.07	.44	.00	.00	.120	192	82	94	70

The original scoring for the six measures were on two different scales, a percentage and an average number scale. In order to arrange them in ascending order of performance using all six measures, the measures were converted to standardized scores where the lowest score on a measure equals zero and the highest score one. In this way the relative magnitude of the scores were maintained while allowing a mean score to be computed across all six measures for each participant to determine their relative performances. The experience measures are shown on the right of the table. The intent of this table is to give a more comprehensive representation of the negative nature of the knowledge structure measures/experience relationship.

To get a feel for the generally negative relationship between the knowledge structure measures and the experience measures, simply scan down the four columns on the right side that contain the experience measures. If the anticipated positive relationship exited, there would be a general decrease in these values as you go down a column. Scanning the columns indicates that this is not the case. In fact, the correlation between the mean score over the six knowledge structure measures with each of the four experience measures is negative. For two of the experience measures, this negative correlation is significant at the .05 level—Months In Service = -.380 and Months In Unit ≥ Battalion = -.302.

The two highlighted rows represent the scores of two individuals that are especially indicative of the negative relationship. The one at the top is the participant with the least amount of time in the Army and the next to least amount of time in the other three experience categories. Yet this individual's mean score on the knowledge structure measures is the highest of the 31 participants. He actually had the highest score on three of the six individual measures. The highlighted row toward the bottom of the table is the participant with the greatest amount of experience in tactical positions, tactical units, and units ≥ battalion. Yet this individual's mean score on the knowledge structure measures is third from lowest of the 31 participants.

If these two were simply exceptions from a generally positive trend, it would still be pretty damning for our hypotheses. But they are extremes of a generally negative trend. It was decided, therefore, to review these two participants' protocols to see if additional cues might be obtained as to why this negative effect occurred.

When comparing the two protocols, the first contrast noted is the extreme difference in response latency. The more experienced individual began answering the response requirements much sooner than the less experienced individual in all three response conditions. It is interesting to note that across all participants, response latencies in the Describe response condition have highly significant ($p<.005$) negative correlations with all four experience measures. The Describe condition is the only one in which the stimulus material is taken away from the participants before they respond. Under this condition, more experienced individuals seem able to "comprehend" the situation much faster than less experienced individuals. This is consistent with the superior "chunking" capability of experts found in many domains (Simon, 1985). Among the experience measures, Months In $\geq$ Battalion, has the greatest number of significant negative relationships with task response latencies. This is indicative of a faster comprehension of these scenarios at battalion and above based on relevant experience.

Concerning the two protocols themselves, the more experienced individual tended to respond more directly to the response requirement. The requirement asked for the relevant characteristics, the relevant attributes, and the relevant relationships. It appeared that the more experienced individual adhered to relevancy and gave just those he considered relevant while the less experienced officer seemed to be trying to come up with all the entities he could associate. The more experienced officer also stuck to just what was on the graphic in the Describe condition, resulting in just Level 1 perceptual responses.

Perhaps the most glaring differences in the two protocols is the amount of elaboration and explanation. The more experienced individual tended to simply state the characteristic, attribute or relationship without explaining or justifying it. The less experienced individual gave fairly lengthy explanations and justifications.

One can see how these response "styles" affect the knowledge structure measures used. The direct responses of the more experienced officer produced lower abstraction and depth scores than the elaborations of the less experienced officer. Even the amount of action in the protocols is affected by the amount of elaboration.

Summary and Conclusions

Only one of the tested hypotheses is supported by the findings. We found no significant relationship between amount of task-related experience and the number of attributes that can be identified, as was stated in hypothesis H2. The measure, Number of Attributes, was significantly affected by the particular scenario being used. But when the scenario effect was eliminated, there still was no significant correlation with the experience measures. A study of the scores does not suggest a ceiling effect as one individual produced 289 attributes and two others had 195 and 145, well above the mean of 77.

The other two tested hypotheses are not supported by the findings. Hypothesis H1 predicts that the amount of interrelationship (integrity) of an officer's battlefield knowledge base increases with relevant experience. The findings suggest the opposite effect. The only significant correlations of the depth and relations measures with experience were all negative and the trend across all such correlations was negative. The exact same was true in relation to hypothesis H3 which predicts that the level of abstract or implicit and inferential task statements made by an officer increases with relevant experience. Here again the only significant effects of experience were negative and the general trend of correlations was negative.

Hypothesis H4, predicts that relatively simple and direct measures of officers' knowledge base integrity and abstraction levels will correlate significantly with the same measures taken in a more complex battlefield problem. This hypothesis was not tested. The data were collected, but the analysis of the complex problem protocols proved beyond the resources of this project. If time and resources permit, it might be tested at a later time.

Why hypotheses H1 and H3 were counterindicated by these results is a matter of speculation. The expertise literature indicates that they do occur in several other domains. There are many possible specific reasons for the generally negative results obtained here, but they all fall into two general categories. Either, in fact, there exists a negative relationship (or at best no relationship) between on-the-job experience of Army officers and growth in knowledge structuring as defined here, or the methods and/or measures used to test knowledge structures in this research are misleading (or at best inadequate).

The first of these two general conclusions seems counterintuitive. The weight of educational, training, and cognitive research evidence and one's personal experience tells us that as we gain experience in a domain we are not only adding new facts but tying facts together. Thus we are better able to relate new experiences to what we have learned in the past, seeing not just similarities (generalizing) but also differences (differentiating). Thus an experienced individual is better able to both classify and define and explain a new situation.

Why did we not see this in our data? If we are to assume that the first general conclusion is true, it would seem to mean that officers who have held tactical positions and worked in tactical units are not receiving sufficient tactical training in these units to build and maintain a superior knowledge base structure compared to those with less of this experience.

Federico (1995) conducted research with naval officers with predictions similar to ours. He found no evidence that expert tactical action officers had better structured or organized knowledge nor better ways of accessing their knowledge than did novices. Nor were the experts more apt to attend to "deeper" more abstract aspects of the situation than novices. Another study of naval officers by Marshall, et al (1996) had a similar finding in that experienced tactical officers appeared to be responding to more track-specific, surface cues than to more abstract, "big picture" cues. It may be that military officers do not have the opportunity to practice their tactical art frequently enough and with enough objective feedback to attain the kind of knowledge structures associated with these qualities. The research that has shown these kind of expert-novice differences involves domains with high frequency of practice and rapid objective feedback such as weather forecasting, radiology and other medical fields, physics, computer programming, etc. However, this remains only speculation, there is better evidence that the fault lies in the method and measures used in this research.

If the first general conclusion were true, it would not explain the negative correlations obtained. We would expect simply no relationship with experience like that obtained in the Federico research. If, in fact, the construct of increasing knowledge structure complexity and inferential capability with experience is valid and Army officers' careers reflect this construct, then our method and measures did not tap it. In fact, our method and measures seem to have actually "hidden" the knowledge bases of experienced officers relative to less experienced officers.

The comparison between the good performer/low experience officer and the poor performer/high experience officer described in the Results section suggests that the experienced officers may be more direct in their responses, producing less protocol than less experienced officers. A comparison of the number of lines of protocol produced with scores on the six knowledge structure measures appearing in Table 5 reveals highly significant relationships ($p<.005$). For example a comparison between the summary measures of average lines of protocol across all three response conditions with average score on the six knowledge structures measures produces a correlation of .573 ($p<.001$). There were significant correlations of protocol size not just with the measures of average depth of development, where it would be naturally expected, but also with the <u>proportional</u> measures of percent of Abstract Level 3 entities and percent of Action entities. Thus the more a participant talks the more apt they are to produce a larger proportion of abstract and action entities along with greater depth in their descriptions. It seems the measures used are highly affected by how verbal the participant is.

It might be argued that the more someone knows, the more they've got to say, but this is not always the case. Ericsson & Simon (1993) have found that experts' protocols are often briefer because of their greater use of recognition and retrieval. The less experienced individual has to "think through" their decisions, actually <u>creating</u> the justifications for them. The expert, on the other hand, has the decision cued directly by the stimulus material by virtue of prior experience with it. Our measures simply asked for the relevant entities, attributes and relations and whatever the participant produced was the product used in the analysis. There was no probing for any further knowledge that might be behind the response. This might well be the reason for our results.

It should be added that there are no significant relationships between protocol size and amount of experience in this experiment. There are consistent negative correlations between protocol size and Time In Service and Time In Unit $\geq$ Battalion, but none are significant. There are, no doubt, many factors that come into play to produce our results. However, it might be worthwhile to test the same or similar measures in an experiment where there is additional immediate probing for unverbalized knowledge.

In conclusion, the method and measures of battlefield knowledge structures tested in this research were not validated in relation to job experience. It remains an open question as to whether further adjustments and refinements to both might yet produce a valid measurement instrument.

References

Anderson, John R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Bellezza, Francis S. (1992). Mnemonics and expert knowledge: Mental cuing. In Robert R. Huffman (Ed.) *The psychology of expertise: Cognitive research and empirical AI.* NY: Springer-Verlag.

Bibby, Peter A. (1992). Mental models, instruction, and internalization. In Y. Rogers, A. Rutherford, and P.A. Bibby (Eds.) *Models of the mind.* London: Academic Press.

Ceci, Stephen J. (1996). *On intelligence: A bioecological treatise on intellectual development.* Cambridge, MA: Harvard University Press.

Ceci, Stephen J. & Ruiz, Ana (1992). The role of general ability in cognitive complexity: A case study of expertise. In Robert R. Huffman (Ed.), *The psychology of expertise: Cognitive research and empirical AI.* NY: Springer-Verlag.

Clancy, Willaim (1988). Acquiring, representing, and evaluating a competence model of diagnostic strategy. In Michelene T.H. Chi, Robert Glaser, and Marshall J. Farr (Eds.) *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Damasio, Antonio R. (1994). *Descartes' error: Emotion, reason, and the human brain.* NY: Avon Books.

diSessa, Andrea A. (1983). Phenomenology and the evolution of intuition. In Dedre Gentner & Albert L. Stevens (Eds.) *Mental models.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Duff, Simon, C. (1992). Mental models as multi-record representations. In Y. Rogers, A. Rutherford and P.A. Bibby (Eds.) *Models of the mind.* London: Academic Press.

Ericsson, Anders K. & Simon, Herbert A. (1993). *Protocol Analysis.* Cambridege, MA: MIT Press.

Federico, Pat-Anthony (1995). Expert and novice recognition of similar situations. *Human Factors, 37,* 105-122.

Fischbach, Gerald D. (1992). Mind and brain. *Scientific American*, September, 48-57.

Fischoff, Baruch (1988). Judgement and decision making. In Robert J. Sternberg and Edward E. Smith (Eds.) *The psychology of human thought.* NY: Cambridge University Press.

Fletcher, K.E. & Huff, A.S. (1990). Argument mapping. In Anne Sigismund Huff (Ed.) *Mapping strategic thought.* NY: John Wiley & Sons.

Forsythe, George B. & Barber, Herbert F. (1992). Military strategic thinking: Expert-novice differences in the structure and content of cognitive representations. Paper given at the American Educational Research Association annual meeting, San Fransico, April, 1992.

Frederiksen, Norman (1986). Toward a broader conception of human intelligence. In Robert J. Sternberg and Richard K. Wagner (Eds.) *Practical intelligence.* NY: Cambridge University Press.

Gardner, Howard (1993). *Multiple intelligences: The theory in practice.* NY: BasicBooks.

Glaser, Robert (1984). Education and thinking: The role of knowledge. *American Psychologist, 39,* 93-104.

Glaser, Robert & Chi, Michelene T.H. (1988). Overview. In Michelene T.H. Chi, Robert Glaser, and Marshall J. Farr (Eds.) *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Goldman-Radic, Patricia S. (1992). Working memory and the mind. *Scientific American,* September, 111-117.

Groen, Guy J. & Patel, Vimla, L. (1988). The relationship between comprehension and reasoning in medical expertise. In Michelene T.H. Chi, Robert Glaser & Marshall J. Farr (Eds.) *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Holyoak, Keith J. & Nisbett, Richard E. (1988). Induction. In Robert J. Sternberg and Edward E. Smith (Eds.) *The psychology of human thought.* NY: Cambridge University Press.

Humphreys, Patrick, Wisudha, Ayleen, Wooler, Stuart, & Oldfiele, Agi (1990). *Handling decision problems: A structuring language and interactive modules.* (Research Note 90-38) Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. AD A223214

Johnson, Eric J. (1988). Expertise and decision under uncertainty: Performance and process. In Michelene T.H. Chi, Robert Glaser, and Marshall J. Farr (Eds.) *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnson-Laird, P.N. (1989). Mental models. In Michael I. Posner (Ed.) *Foundations of cognitive science.* Cambridge, MA: The MIT Press.

Klein, Gary A., Thordsen, Marvin L. & Claderwood, Roberta (1990). *Descriptive models of military decision making.* (Research Note 90-93) Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. AD A226884

Lawrence, Jeanette A. (1988). Expertise on the bench: Modeling magistrates' judicial decision-making. In Michelene T.H. Chi, Robert Glaser, and Marshall J. Farr (Eds.) *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Lesgold, Alan (1988). Problem solving. In Robert J. Sternberg & Edward E. Smith (Eds.) *The psychology of human thought.* NY: Cambridge University Press.

Lesgold, Alan, Glaser, Robert, Rubinson, Harriet, Klopfer, Dale, Feltovich, Paul & Wang, Yen (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In Michelene T.H. Chi, Robert Glaser and Marshall J. Farr (Eds.) *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Marshall, Sandra P., Christensen, Scott E. & McAllister, James A. (1996). Cognitive differences in tactical decision making. In *1996 Command and Control Research & Technology Symposium, Command and Control in the Information Age*, National Defense University Publication.

Miller, George A. (1985). Trends and debates in cognitive psychology. In A.M. Aitkenhead and J.M. Slack (Eds.) *Issues in cognitive modeling.* London: Lawrence Erlbaum Associates.

Norman, Donald A. (1985). Twelve issues for cognitive science. In A.M. Aitkenhead & J.M. Slack (Eds.) *Issues in cognitive modeling.* London: Lawrence Erlbaum Associates.

Posner, Michael I. (1988). Introduction: What is it to be an expert? In Michelene T.H. Chi, Robert Glaser and Marshall J. Farr (Eds.) *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Reimann, P. & Chi, M.T.H. (1989). Human expertise. In K.J. Gilhooly (Ed.) *Human and machine problem solving.* London: Plenum Press.

Rochester, S.R., Thurston, Sharon & Rupp, Judith (1977). Hesitations as clues to failures in coherence: A study of the thought-disordered speaker. In Sheldon Rosenberg (Ed.) *Sentence production.* NY: Lawrence Erlbaum Associates.

Royer, James M., Cisero, Cheryl A, & Carlo, Maria S. (1993). Techniques and procedures for assessing cognitive skills. *Review of Eduction Research, 63,* 201-243.

Rumelhart, Daivid E. (1989). The architecture of mind: A connectionist approach. In Michael I. Posner (Ed.) *Foundations of cognitive science.* Cambridge, MA: The MIT Press.

Rumelhart, David E. & Norman, Donald A. (1985). Representation of knowledge. In A.M. Aitkenhead & J.M. Slack (Eds.) *Issues in cognitive modeling.* London: Lawrence Erlbaum Associates.

Schank, Roger C. (1985). Reminding and memory organization. In A.M. Aitkenhead & J.M. Slack (Eds.) *Issues in cognitive modeling.* London: Lawrence Erlbaum Associates.

Scribner, Sylia (1986). Thinking in action: Some characteristics of practical thought. In Robert J. Sternberg & Richard K. Wagner (Eds.) *Practical Intelligence.* NY: Cambridge University Press.

Simon, Herbert A. (1985). Information processing theory of human problem solving. In A.M. Aitkenhead & J.M. Slack (Eds.) *Issues in cognitive modeling.* London: Lawrence Erlbaum Associates.

Smith, Edward E. (1989). Concepts and induction. In Michael I. Posner (Ed.) *Foundations of cognitive science.* Cambridge, MA: The MIT Press.

Solick, Robert E., Spiegel, Douglas K., Lussier, James W. & Keene, S. Delane (1997). *Visualization and judgmental forecasting of simulated battles.* (Research Note 97-08) Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. AD A328489

Soloway, Elliot, Adelson, Beth & Ehrlich, Kate (1988). Knowledge and processes in the comprehension of computer programs. In Michelene T.H. Chi, Robert Glaser & Marshall J. Farr (Eds.) *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Sternberg, Robert J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence.* NY: Cambridge University Press.

Sternberg, Robert J. & Frensch, Peter A. (1992). On being an expert: A cost-benefit analysis. In Robert R. Huffman (Ed.) *The psychology of expertise: Cognitive research and empirical AI.* NY: Springer-Verlag.

Staszewski, James J. (1988). Skilled memory and expert mental calculation. In Michelene T.H. Chi, Robert Glaser and Marshall J. Farr (Eds.) *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates.

VanLehn, Kurt (1989). Problem solving and cognitive skill acquisition. In Michael I. Posner (Ed.) *Foundations of cognitive science.* Cambridge, MA: The MIT Press.

Voss, James F. & Post, Timothy A. (1988). On the solving of ill-structured problems. In Michelene T.H. Chi, Robert Glaser, & Marshall J. Farr (Eds.) *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Wagner, Richard (1986). The search for intraterrestrial intelligence. In Robert J. Sternberg and Richard K. Wagner (Eds.) *Practical intelligence.* NY: Cambridge University Press.

Wiser, Marianne & Carey, Susan (1983). When heat and temperature were one. In Dedre Gentner & Albert L. Stevens (Eds.) *Mental models.* Hillsdale, NJ: Lawrence Erlbaum Associates.