INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Mental models of physical systems: Examining the relationship between knowing and doing

Rowe, Anna L., Ph.D.

Rice University, 1994
RICE UNIVERSITY

MENTAL MODELS OF PHYSICAL SYSTEMS: EXAMINING THE RELATIONSHIP BETWEEN KNOWING AND DOING

by

ANNA L. ROWE

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE DOCTOR OF PHILOSOPHY

APPROVED, THESIS COMMITTEE

David Lane, Associate Professor, Chair
Department of Psychology

Michael Watkins, Professor
Department of Psychology

John Brelsford, Professor
Department of Psychology

Richard Grandy, Professor
Department of Philosophy

Nancy Cooke, Assistant Professor
New Mexico State University

Houston, Texas
May, 1994
Mental Models of Physical Systems:
Examining the Relationship Between Knowing and Doing

by

Anna L. Rowe

ABSTRACT

Although use of the mental model construct has proliferated in recent applied research, the construct lacks an agreed upon method of measurement. Importantly, the validities of the different measurement techniques have not been established. The purpose of this research was to evaluate the validity of several mental model measurement methods, with the criterion being task performance. Additionally, mental models were examined in the context of a real-world, complex problem-solving situation. Four mental model measurement techniques were evaluated: a laddering structured interview, concept relatedness ratings, a diagramming structured interview, and think aloud while troubleshooting. Nineteen U. S. Air Force technicians varying in troubleshooting expertise each completed the mental model measures. In addition, the technicians each worked to troubleshoot a moderately difficult problem. The results indicate that two of the evaluated techniques were each independently predictive of troubleshooting performance: the laddering structured interview and the concept relatedness ratings. Because these measures are predictive of performance, it is recommended that researchers consider using these techniques in work requiring the measurement of mental models. The next step in this line of research involves a characterization of the kinds of information offered by different mental model measures.
This research also revealed basic psychological issues regarding the development of expertise. Specifically, the results suggest that expertise may not develop in a monotonic fashion. Changes in knowledge may not be adequately represented as simple monotonic increases in similarity to some ideal knowledge representation. The results also indicated that the provision of a context when measuring mental model knowledge with different methods may produce conflicting results. Finally, the benefits associated with using performance as the criterion when assessing validity are discussed.
Acknowledgments

I would like to thank Nancy Cooke for the advise, support, and friendship she has given me during the past several years. Nancy provided invaluable direction and insight throughout the conduct of this project. Special thanks are extended to David Lane who willingly took over the role of committee chair upon Nancy's departure from Rice. I would also like to thank my committee members, Mike Watkins, John Brelsford, and Richard Grandy, for their assistance with this work.

I am grateful for the support provided by the U.S. Air Force and particularly the Basic Job Skills Research group. Many Air Force personnel played various roles in this research endeavor, and several people deserve special attention, namely Ellen Hall, TSgt Ron Kane, and TSgt Chris Kruse. I sincerely appreciate the assistance they each provided.

My greatest expression of thanks goes to my family. I could not have attempted graduate school as a single parent without the constant support and encouragement provided by my own parents, Pat and Linda Rowe. Thanks Mom and Dad! I am also grateful for the support I received from all of my family and friends, particularly Julie, Will, Kristie, Tracy, Tammy, and Alan. Finally, Dacia Sil Hallbert, the person most affected by my academic endeavors, is deserving of extra attention. I would like to dedicate this work to her.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Accuracy and Time Measures</td>
<td>5</td>
</tr>
<tr>
<td>Interviews</td>
<td>8</td>
</tr>
<tr>
<td>Process Tracing</td>
<td>9</td>
</tr>
<tr>
<td>Structural Analysis</td>
<td>10</td>
</tr>
<tr>
<td>Method</td>
<td>15</td>
</tr>
<tr>
<td>Problem Selection</td>
<td>15</td>
</tr>
<tr>
<td>The RWR Wiring--Broken Wire Problem</td>
<td>23</td>
</tr>
<tr>
<td>Subjects</td>
<td>27</td>
</tr>
<tr>
<td>Materials and Procedure</td>
<td>28</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>34</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>37</td>
</tr>
<tr>
<td>Laddering Structured Interview</td>
<td>37</td>
</tr>
<tr>
<td>Relatedness Ratings</td>
<td>42</td>
</tr>
<tr>
<td>Diagramming</td>
<td>46</td>
</tr>
<tr>
<td>Think Aloud</td>
<td>50</td>
</tr>
<tr>
<td>Questionnaires</td>
<td>54</td>
</tr>
<tr>
<td>General Discussion</td>
<td>55</td>
</tr>
<tr>
<td>Summary and Recommendations for Measuring Mental Models</td>
<td>55</td>
</tr>
<tr>
<td>Findings on Measuring Mental Models</td>
<td>60</td>
</tr>
<tr>
<td>Conclusions</td>
<td>67</td>
</tr>
<tr>
<td>References</td>
<td>70</td>
</tr>
<tr>
<td>Appendix</td>
<td>Page</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
</tr>
<tr>
<td>Appendix A</td>
<td>80</td>
</tr>
<tr>
<td>Appendix B</td>
<td>86</td>
</tr>
<tr>
<td>Appendix C</td>
<td>92</td>
</tr>
<tr>
<td>Appendix D</td>
<td>97</td>
</tr>
</tbody>
</table>
List of Tables

Table 1 ........................................................................................................ 6
Table 2 ........................................................................................................ 21
Table 3 ........................................................................................................ 22
Table 4 ........................................................................................................ 25
Table 5 ........................................................................................................ 27
Table 6 ........................................................................................................ 38
Table 7 ........................................................................................................ 39
Table 8 ........................................................................................................ 40
Table 9 ........................................................................................................ 43
Table 10 ...................................................................................................... 44
Table 11 ...................................................................................................... 47
Table 12 ...................................................................................................... 52
Table 13 ...................................................................................................... 56
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>26</td>
</tr>
<tr>
<td>Figure 2</td>
<td>27</td>
</tr>
<tr>
<td>Figure 3</td>
<td>31</td>
</tr>
<tr>
<td>Figure 4</td>
<td>36</td>
</tr>
<tr>
<td>Figure 5</td>
<td>37</td>
</tr>
<tr>
<td>Figure 6</td>
<td>45</td>
</tr>
<tr>
<td>Figure 7</td>
<td>49</td>
</tr>
</tbody>
</table>
INTRODUCTION

Recent research suggests that understanding a complex physical system and successfully interacting with it requires several different types of knowledge including: knowledge of the basic system components, the possible states of those components, and how those components are interrelated (Hegarty, 1991). Such knowledge forms an internal mental representation or "mental model" of the system (Gentner & Stevens, 1983; Staggers & Norcio, 1993). According to Norman (1983) mental models are acquired through interactions with the system, and possession of a mental model provides the system user with predictive and explanatory power. Although, the mental model construct has also been used by researchers interested in language comprehension and inference (Johnson-Laird, 1983), this paper is concerned with people's mental models of complex physical systems, what Brewer (1987) refers to as causal mental models.

The idea that humans make use of internal mental models in their interactions with complex systems is not new. In 1943 Craik concluded that if a system user held an internal "small-scale model" of external reality and of its possible actions, the user would be capable of acting in a much fuller, safer, and more competent manner. According to Rasmussen and Jensen (1974), a user operating a complex, technical system utilizes a mental model representation of the system to process information that the system presents. Rasmussen and Jensen further state that, "the presence of a mental model is taken as given, since meaningful data-processing basically has to be based upon a representation of the constraints which a system impose upon the interrelation of the data it presents" (p. 293).
The mental model construct makes intuitive sense, and use of the construct for pragmatic purposes such as system design and instruction improvement has been promoted (Rouse & Morris, 1986). In fact, the mental model construct has been used in a variety of research settings ranging from examinations of software design (e.g., Carroll & Olson, 1983), to interface design (e.g., Gillan, Breedin & Cooke, 1992; Kellogg & Breen, 1991), and user training (e.g., Carroll & Olson, 1987; Halasz & Moran, 1983). It should be noted however, that some researchers (e.g., Kieras, 1988) hold that mental models are not always necessary for successful interactions with complex devices. Kieras (1988) offers using the ordinary telephone system to make a long distance telephone call as an example of a case in which a complex system can be successfully used without invoking a mental model of the system. Regardless, use of the mental model construct in applied settings has proliferated.

Despite its popularity, the mental model construct lacks an explicit, agreed upon definition (Rouse & Morris, 1986; Wilson & Rutherford, 1989). As Lansdale (1985) stated, many authors "emphasize the significance of 'models' or 'user's representations' without feeling the need to specify what they mean by those terms" (p. 245). Occasionally, authors do not define the construct when they use it in their research (e.g., Nardi & Zarmer, 1991), but more often the available definition is vaguely specified, consisting of imprecise terms. For example, a mental model for a calculator or computer has been defined as "the user's conception of the 'invisible' information processing states and transformations that occur between input and output" (Bayman & Mayer, 1984, p. 189). However, the authors do not specify what is meant by "invisible information processing states and transformations." In addition, many of the specified definitions have been given by researchers who have different
orientations, reducing the compatibility of the definitions. For example, similar terminology has very different meaning to members of different disciplines (e.g., the term model), creating a communication barrier (Wilson & Rutherford, 1989). Finally, the difference between a mental model as knowledge and a mental model as information is not well clarified. For example, several researchers (e.g., Bayman & Mayer, 1984; Frese et al., 1988; Gentner & Gentner, 1983; Halasz & Moran, 1983; Sein & Bostrom, 1989) have presented subjects with information regarding the functioning of a system and subsequently assumed that the subjects held the taught system model. In this type of situation, the authors often fail to take into account subjects' interpretations of the presented information which may serve to alter the information. Despite these problems, the different definitions do have a degree of commonality. However, the commonality is greater in defining what a mental model achieves, rather than in defining what a mental model is (Lansdale, 1985). Specifically, possession of a mental model is typically assumed to guide actions and facilitate interpretations of system behavior (Young, 1981).

Although the mental model construct lacks a clear definition, research employing it has continued, with different researchers using their own operationalizations of the construct. This lack of a clear measurement approach has led to increased confusion about the exact nature of the mental model construct. The body or network of evidence surrounding a construct has been termed a "nomological network" (Chronbach & Meehl, 1955). Ideally, research adds to and strengthens this network. However, "nomological noise" (Messick, 1981), or confusion in the network, can arise from several sources, one of which is the situation in which researchers each use unique operationalizations to elaborate the construct. Such a situation is problematic because no one
measure is a pure exemplar of the construct in question; the measure also contains variance due to other constructs and method contaminants. For example, if two different researchers measure one person's mental model of a complex system with two different methods, the methods may tap somewhat different types of knowledge, making generalizations across the two studies impossible even though both researchers are purporting to measure the same construct from the same person. Furthermore, confusion (i.e., nomological noise) would result from an assimilation of the results of these measures into the body of evidence (i.e., the nomological network) surrounding the construct.

The primary purpose of measurement in science is to enable the expression of the functional relations between constructs in terms of mathematical equations. Development of a scientific theory is virtually impossible unless the variables involved in the theory can be adequately measured (i.e., with reliable and valid instruments). Thus, what is required is an assessment of the different methods of measuring mental models. This research assesses the validity of several different mental model measurement methods. There is no "true" mental model to serve as a criterion, making only indirect measures of validity possible (Cooke, 1994a). Because performance enhancement is often the goal in applied settings, this research took a pragmatic approach to assessing validity by examining how well the methods measure knowledge that is relevant to performance. A "good" mental model measure should provide output that reflects knowledge differences. These knowledge differences should then coincide with performance differences.

Given Kieras' (1988) assertion that mental models are not always necessary for successful performance, an ill-structured problem solving domain was chosen which appears to promote the use of mental models: avionics
troubleshooting (Gott, 1989; Hall, Gott, & Pokorny, 1990; Rasmussen & Jensen, 1974). Specifically, the mental models of subjects tasked with diagnosing faults in complex avionics systems were assessed. What follows is a review of the various methods which have been utilized to measure mental models. The different methods of measuring mental models may be classified into four categories: 1) accuracy and time measures, 2) interviews, 3) process tracing/protocol analysis, and 4) structural analysis (see Table 1).

**Accuracy and Time Measures**

When accuracy and time measures are used to assess mental models, subjects are given a set of problems to solve, and problem-solving behavior is examined to infer the presence or absence of a mental model. Thus, accuracy and time measures offer only indirect evidence for the existence of mental models (Staggers & Norcio, 1993). Researchers have used a wide variety of problem-solving topics when making inferences about mental models. For instance, researchers interested in mental models of motion have examined error patterns in subjects' predictions about an object's motion (McCloskey, 1983) and in undergraduate students' physics lab reports (Clement, 1983). Similarly, Gentner and Gentner (1983, Experiment 1) examined the relationship between subjects' electronic circuitry problem-solving performance and their self-reported use of a particular analogy for the function of electricity.

Researchers have also used problem-solving performance to examine the effect of having a particular mental model after teaching different mental models of a system to different groups of subjects. For example, researchers have compared subjects' calculator problem-solving performance following instruction on different mental models of the calculator's functioning (Bayman & Mayer, 1984; Halasz & Moran, 1983). Likewise, after teaching one of three
Table 1
*Measures of mental model knowledge.*

---

**Accuracy and time measures**

**Interviews**
- Unstructured interview
  - "how does it work?" questions
- Structured interview
  - focus on...
    - ...a specific module
      - Ask Where is it located? What is it? What does it do? How does it work?
    - ...diagrams
      - -enumerate concepts
      - -show physical relations
      - -show functional relations (functional flow analysis)
      - -encircle related components
      - -sort for structural modeling
    - ...explanation of system behavior

**Process Tracing**
- Verbal Reports/Think Aloud
  - on-line
  - off-line
- Nonverbal data
  - (e.g., actions, eye movements)

**Structural Modeling**
- Multidimensional scaling
  - Metric
  - Non-metric
- Discrete feature models
  - Hierarchical clustering
  - Additive trees
  - Extended trees
  - Additive clustering
  - Pathfinder network scaling
- Direct elicitation of structure
  - Free association
  - Graph construction
  - Question-answering
  - Drawing Closed Curves
different models of electronic circuitry to their subjects, Gentner and Gentner (1983, Experiment 2) compared subjects' problem solving performance to determine if possession of a particular model of electricity led to superior performance on problems which "fit" the model. Kieras and Bovair (1984) compared performance in groups given different training in the use of a control panel device by examining actions taken and response latency. Similarly, Sein and Bostrom (1989)'s study of instruction programs utilizing different conceptual models included an examination subjects' performance on a set of hands-on, near- and far-transfer tasks in an electronic mail filing system.

Researchers have also examined various performance variables to evaluate the effects of mental model versus no mental model training programs. For example, Frese et al. (1988) compared the performance of subjects who had received one of three different training programs for a word processing system: 1) a sequential program which did not encourage the active development of a mental model, 2) a hierarchical program which provided an explicit and integrated mental model of the presented system, and 3) an active, exploratory program which promoted user mental model development through the development of hypotheses regarding the functioning of the system. Performance measures included free recall of commands, a corrected error score from a speed typing task, an inefficiency score based on a text-editing task, and a transfer score based on a task requiring the use of a new command, among others.

In summary, using accuracy and time measures is a popular approach for making inferences about mental models. When this technique is used, researchers are able to observe and record actual performance, offering the opportunity to pinpoint problem-solving strategies which may not be
consciously accessible. Although this approach has intuitive appeal and face
validity, the data gathered from this technique are rather impoverished in that
the cognitive underpinnings of actions are not elaborated. Instead, the actions
must be examined in an attempt to uncover action patterns (or errorful actions)
which are indicative of the presence (or absence) of a particular model. Thus,
action data offer only indirect evidence (Staggers & Norcio, 1993). In addition,
recent research indicates that time and errors do not always map neatly onto a
single mental model (Cooke & Breedin, in press). As a consequence, most
attempts to measure mental models in the literature have combined this basic
measure with one or more of the relatively richer methods described below.

**Interviews**

The results of an interview may be examined to infer the presence of a
mental model. Interviews to elicit mental models run along a continuum of
structure. When a researcher conducts an unstructured interview neither the
content nor the course of the interview are established prior to the interview. In
general, unstructured interviews begin with a “tell me how it works” question.
From this point both the content and course of the interview evolve as the
interview progresses. For example, unstructured interviews have been used to
examine explanations given for simple phenomena in physics (DiSessa, 1983)
and to infer theories of home heat control (Kempton, 1986). Unstructured
interviews may also be conducted after the subject has solved a problem at
which point the researcher asks the subject about his/her problem-solving
performance (e.g., McCloskey, 1983).

Structured interviews, on the other hand, follow some sort of prespecified
format. The structured interview may focus on one or more topic. For instance,
subjects may be asked to discuss the functioning of a particular complex system
as in Payne's (1991) descriptive study of subjects' mental models of automatic
teller machines (ATMs) and in Staggers and Norcio's (1993) study of nurses'
mental models of SPSSX, a statistics package. Similarly, Vosniadou and
Brewer (1992) explored children's mental models of the earth through a series
of questions about the shape of the earth. Some researchers confine subjects'
discussions to aspects of specific system components. For example, Gitomer
(1984) asked subjects to discuss a specific system component in terms of its
location, purpose, and/or function. Subjects may also be asked to create
diagrams of system components to enumerate concepts, designate related
components, and show physical and/or functional relations (Gray, 1990;
Gitomer, 1984; Hall, Gott, & Pokorny, 1990). Finally, the structured interview
may revolve around a discussion of specific examples of system behavior

In summary, interviews offer a relatively natural and uncontrived setting for
the measurement of mental models. In addition, the data resulting from
interviews offer a richer representation of the cognitive processes underlying
performance. This richer data set does not come without a price however.
Interviews make use of verbal data which may prove limiting, and interview data
sets are usually rather large and difficult to interpret.

Process Tracing

When process tracing is used to measure mental models, subjects are
presented with a problem and are asked to think aloud as they solve the
problem. The protocol is subsequently analyzed to make inferences about
mental models. Although such verbal reports have been criticized on the
grounds of their accuracy and completeness (e.g., Nisbett & Wilson, 1977),
others (Ericsson & Simon, 1984) have attempted to define the conditions under
which verbal protocols are appropriate. Nonetheless, this method of measuring mental models is very popular. For example, think-aloud protocols have been used to examine electronics troubleshooting (Rasmussen & Jensen, 1974) and physics problem solving (Clement, 1983; Larkin, 1983). Such protocols have also been used to determine the influence of various problem representations on problem solving (Greeno, 1983). Halasz and Moran (1983) analyzed subjects' think aloud protocols as they solved problems using a stack calculator to infer how users employed mental models in solving novel problems. Similarly, Gray (1990) used think-aloud protocols gathered during subjects' interactions with a hypertext system to investigate mental model development during hypertext navigation. Researchers have also analyzed verbal protocol data to determine if various developed models would predict performance (Williams, Hollan, & Stevens, 1983).

In summary, process tracing allows the exploration of the cognitive structure and processes underlying task performance, and the data collection stage is relatively easy to carry out. However, the use of process tracing may be limiting because the technique makes use of verbal data. Although verbal protocols provide evidence about what a subject is thinking about, they do not necessarily provide information on how that subject thinks. In addition, process tracing results in large, rather unorderly data sets that are qualitative in nature, making meaningful interpretations difficult.

**Structural Analysis.**

When structural analysis is used to measure mental models, pairwise proximity estimates for a set of items are gathered. These estimates are then submitted to a descriptive multivariate statistical technique which reduces the estimates to a simpler, mathematically derived organization which can then be
evaluated. Examples of structural analysis techniques include multidimensional scaling (MDS), cluster analysis, and network scaling techniques such as Pathfinder (Schvaneveldt, 1990).

Researchers have used such structural techniques to compare the mental models of different groups of subjects. For instance, Gillan, Breedin, and Cooke (1992) used MDS and Pathfinder to compare three groups of subjects (software experts, human factors experts, and a control group of computer users with no design experience) in terms of their mental models of interface design concepts. Likewise, Gillan and Breedin (1990) used hierarchical cluster analysis and Pathfinder to compare the mental models of the human-computer interface held by human factors experts and software experts. Gitomer (1984) used cluster analysis and MDS to assess the mental models of an antenna system held by subjects varying in expertise. Researchers have also used structural techniques to assess the similarity between subjects' mental models and some ideal model. For example, Kellogg and Breen (1990) used Pathfinder to derive and compare users' mental models to an idealized system model. Finally, structural techniques have been used as means to determine whether subjects exposed to a particular training program acquire a desired mental model. For example, Coury, Weiland, and Cuoqlock-Knopp (1992) used MDS in their investigation of the mental models of subjects trained to identify the states of an hypothetical system. These researchers also used the MDS representations to evaluate the effect of different display formats.

In summary, structural analytic techniques provide another approach to determining the cognitive underpinnings of performance. The researcher using structural analysis is able to objectively interpret large data sets through the use of various data reduction techniques which may reveal the most meaningful
features of the data. Furthermore, structural techniques do not rely on detailed verbal reports. However, structural techniques may not adequately measure heuristics, rules, or strategies, and the measurement setting may appear artificial and contrived.

Overall, four very different categories of measurement methods have been used in research on mental models, including accuracy and time measures, interviews, process tracing, and structural analysis. In general, each of the different methods has advantages and disadvantages (Cooke, 1994a). Specifically, inferences based on accuracy and time measures are based on actual performance and thus are presumably relevant to that performance. However, these measures are rather impoverished in that the cognitive processes underlying task performance are not elaborated. Interviews and process tracing both offer a means to explore these cognitive processes in a relatively natural and uncontrived setting. However, both techniques often result in large data sets which may be difficult to interpret meaningfully, and their reliance on verbal data may prove limiting. Conversely, structural analytic techniques provide an objective approach to interpreting large amounts of data through data reduction techniques which may reveal the most meaningful features of the data. However, the structural techniques may not adequately measure heuristics, rules, or strategies, and the measurement setting may not appear natural.

Obviously, no one method of measuring mental models has received universal acceptance. Little research has been concerned with an evaluation of the different measurement approaches in terms of their respective reliabilities and validities. Preliminary research has shown that several of the techniques reliably assess the mental models of experts over a period of one week but are
less reliable in assessing the mental models of novices over the same period of time (Rowe, Cooke, Neville, & Schacherer, 1992). Furthermore, the results of other research assessing the converging validity of several structural techniques indicate that the methods converge on the same information to a greater extent than would be expected by chance (Gammack, 1990). Although these steps toward the evaluation of the measurement methods are promising, the relationship between the information provided by the various measures and subsequent performance remains to be delineated.

Thus, the purpose of this research was to evaluate the validity of several mental model measurement methods, with the criterion being actual performance. In other words, the validity of the techniques was examined by determining how well the knowledge represented by each of the techniques related to task performance. Additionally, because most of the research using the mental model construct has involved either abstract and academic subjects such as physics, geometry, etc. or relatively simple versions of "real-world" problem solving such as calculator keystrokes, the current research examined mental models in the context of a real-world, complex problem-solving situation. This research examined the mental models of technicians in the F-15 flight line avionics career field in the United States Air Force. These technicians complete two major types of tasks: 1) identifying and isolating malfunctions of airborne avionics equipment (line replaceable units or LRUs) and repairing them, and 2) verifying, calibrating, modifying, and repairing the test equipment used to test the LRUs (Gitomer, 1984). This research focused on the knowledge needed for successful troubleshooting. Previous research has indicated that mental models are a vital component in successfully troubleshooting ill-defined
problems (Gitomer, 1984; Hall, Gott, & Pokorny, 1990; Rasmussen & Jensen, 1974).

The current research differs from other studies which used multiple methods for measuring mental models (e.g., Gray, 1990; Gitomer, 1984; McCloskey, 1983) in that the primary interest of this research is a comparison of the measurement techniques themselves. The primary interest in previous research endeavors employing multiple measurement techniques has been the assessment of underlying cognitive characteristics (including, but not limited to, mental models), using converging operations. This research also addressed several issues which have been raised in the conduct of previous work in the avionics troubleshooting domain, including the identification of a performance measure. This issue and others were addressed by (1) utilizing the measurement techniques within a specific troubleshooting context, (2) employing cognitively rich mental model measurement techniques, including pairwise comparisons, Pathfinder, and think-aloud reports, that have shown promising results in other domains, (3) using a performance-based criterion (i.e., verbal troubleshooting score), and (4) measuring performance on a continuous scale rather than a dichotomous scale in an attempt to increase measurement sensitivity.

The research consists of two general phases: a problem selection phase and a mental model measurement technique comparison phase. During the problem selection phase, a moderately difficult troubleshooting problem was selected. Problem selection was vital because all experimental materials revolved around one problem. During the measurement technique comparison phase, subjects completed each of the evaluated mental model measures and worked to verbally troubleshoot the selected problem.
METHOD

Problem Selection

All experimental materials were developed in the context of a particular troubleshooting problem in the F-15 flightline avionics communications, navigation, and electronic warfare systems (or C Shop) career field of the U.S. Air Force. The procedure used to select this problem was designed for the selection of a moderately difficult troubleshooting problem. A moderately difficult troubleshooting problem presumably requires the invocation of a mental model for successful troubleshooting. Such a problem should distinguish expert from novice technicians. Specifically, if a problem could be solved through the utilization of a memorized procedure, the problem would be too easy. Novice technicians who had memorized the particular set of procedures would not be distinguishable from experts. On the other hand, if a problem could not be solved by any technician, novice or expert, the problem would not distinguish expert and novice performers due to its high difficulty level. This section describes the procedure used to select a moderately difficult problem. Much of the data used in the selection of this problem were gathered during Stages I-VIII of a PARI cognitive task analysis (Hall, Gott, & Pokorny, 1990) conducted by an Air Force research team in the C Shop career field (Hall, Pokorny, & Kane, 1994) at Eglin Air Force Base. The research team included three psychologists (including the author) and two C Shop experts.

PARI (Precursor, Action, Result, Interpretation) is a cognitive task analysis methodology used by the Air Force as an integrated skill analysis/instructional development tool. The PARI data collection procedure consists of nine stages. In general, the first four stages serve to identify a sample of subject matter experts. These experts then assist the research team in identifying the general
problem solving tasks encountered in the career field and the cognitive skills associated with successfully solving these tasks. The final five stages of PARI involve the development of problem-solving scenarios and the collection of problem-solving interview data from experts and novices as well as a set of follow-up reviews of the data. Hall, Pokorny, and Kane's (1994) PARI data were used in the selection of the moderately difficult troubleshooting problem. Data from Stage IX of this PARI analysis were not pertinent in the selection of a moderately difficult troubleshooting problem and thus were not utilized. The following paragraphs offer only a brief description of the PARI methodology; a more complete account of PARI can be found in Hall, Gott, and Pokorny (1990).

**PARI-Stage I.** The first stage of PARI is designed to identify subject matter experts who then go on to participate in the remaining PARI stages. In order to identify C Shop experts, the Air Force research team conducted individual discussion sessions with technicians who had been identified as the most highly skilled by shop supervisors and were available for participation in the discussions. During the discussion session, the technician was asked to iteratively break down F-15 avionics equipment systems in terms of their component parts. First, the technician identified the subsystems of a particular system (e.g., the Radar Warning Receiver or RWR system is part of the Tactical Electronic Warfare System or TEWS). The technician then broke the subsystems down to the component level, identifying the function of each component as it was named. This break-down continued until the technician believed that the components at the lowest level could not be further subdivided. In addition to iteratively breaking down the equipment systems, the technician addressed job training problems associated with the particular system.
Based on these discussions, the researchers, as a group, determined which of the sampled technicians qualified as experts. This determination was based on three aspects of each technician's discussion: 1) the quality of the verbalized equipment representations, 2) the identification of specific equipment component relations, and 3) the level of clarity in the technician's equipment descriptions. Selection was also based on availability to participate in the PARI sessions. Following the application of this process, the research team designated two of the sampled technicians as subject matter experts: an Air Force Technical Sergeant and an Air Force civil servant. These experts assisted the research team throughout the remaining stages of the cognitive task analysis.

**PARI-Stage II.** The second stage of PARI is designed to establish the training foci associated with the job in question. (PARI was developed to assist in the development of training that targets complex problems.) To form the C Shop training foci, the two experts worked to list and discuss the maintenance tasks (i.e., troubleshooting problems) they felt were difficult. These discussions were facilitated by an exhaustive listing of maintenance tasks for that career field provided by occupational surveys and Air Force Specialty Training Standards. The Air Force research team used two related criteria to classify tasks as cognitively complex: the degree of decision-making required in performing the task and the stability of the task (or system) environment in which problem solving occurs. Tasks were considered cognitively complex if they required decision making (i.e., a procedure specifying step-by-step actions for solving the problem is not available) and if they occurred in an unstable environment (i.e., many factors must be considered in making decisions). Tasks meeting these two criteria were subsequently used by the research team to
facilitate discussion with the two experts. The experts were asked to identify maintenance tasks associated with their jobs which were cognitively demanding and to discuss their reasoning for this assertion. The research team and the experts then decided together whether or not to categorize the task as cognitively complex. Thus, the research team and the experts worked together to identify the cognitively complex tasks associated with the C Shop job.

**PARI-Stage III.** The purpose of the third stage of PARI is the generation and consolidation of the problem types encountered by technicians working on the job. Using the cognitively complex tasks identified in Stage II, the experts worked to generate an exhaustive list of the equipment malfunctions (and their causes) that could initiate troubleshooting for these tasks. The experts independently specified fault instances in cause and effect language (e.g., "bad stimulus routing caused by a stuck relay") for each of the defined tasks. The experts then worked together to consolidate the identified system causes and effects into meaningful categories (e.g., wiring faults). Fault instances were grouped together if they demanded similar knowledge and skills for solution. This grouping resulted in the following typology of problems which could initiate troubleshooting in the C Shop: set up procedure faults, switchology faults, cable faults, wiring faults, and electrical/component faults.

**PARI-Stage IV.** The fourth stage of PARI involves the development of representative troubleshooting problems for each of the problem categories specified in the problem typology. The typology serves to ensure that representative examples of all problem types are generated. The experts individually developed a representative troubleshooting problem for each of the five problem categories. For each of these problems the experts: (1) developed an overview or problem description which listed the fault location and the
symptoms associated with the fault, (2) generated a problem statement that listed the system conditions and symptoms for presentation to other individuals for troubleshooting, and (3) listed the supporting technical documentation (e.g., test procedures, schematics) that would be required by others troubleshooting the problem. At the conclusion of PARI-Stage IV, each of the experts had designed one troubleshooting problem for each problem category.

**PARI-Stage V.** The purpose of the fifth stage of PARI is an anticipation of the supporting information (e.g., Technical Orders or T.O.s) technicians would require to solve the developed problems. In addition, this stage also provides an opportunity for the experts to specify exactly how the various pieces of equipment would function under the faulty conditions. To obtain these sorts of information, the experts worked individually to generate their own solutions to the problems they had developed. The PARI problem-solving structure was used to guide solution generation: Actions, Precursors to actions, Results, and Interpretations of results were elicited. Following the recording of the initial solution, five "rehashes" were conducted in which the expert worked to (1) verify the initial solution, (2) generate alternative results and result interpretations for each step in the solution, (3) identify and evaluate alternative actions for each step in the solution, (4) name alternative appropriate equipment targets (precursors), given the previously executed steps, and (5) group the actions that seem to go together and to explain the basis for these groupings. At the conclusion of PARI-Stage V, the experts had each generated one solution for each of their respective problems.

**PARI-Stage VI.** During the sixth stage of PARI, experts naive to the problems are asked to generate solutions to those problems. Thus, two additional technicians were asked to produce solutions to the generated
problems. These technicians had been identified by the two experts as skilled technicians, and both were at the 7-Skill Level. The Air Force uses a four level classification system to designate skill level: all technicians start at the 3-level and move from the 5- to the 7-level after passing certain training criteria (e.g., demonstrating proficiency running an operational checkout of a particular equipment system). Technicians at the 7-level have reached the highest level of technical proficiency. They then move on to the 9-level which designates them as qualified on management-related tasks. The 9-level designation is attainable only by Senior and Chief Master Sergeants, who no longer have hands-on maintenance responsibilities.

The two technicians worked individually with one of the experts to solve the generated troubleshooting problems. The experts presented only the problems they had developed. Thus, the experts worked individually with each of the technicians. The expert began by presenting the problem statement to the technician (e.g., "During debrief, the crew chief reports an ASP 44"). The technician then worked to isolate the fault and repair the equipment through a series of iterative action-result steps. In each step the technician specified an action and the reason for taking that particular action. The expert responded by informing the technician of the action's effect on the equipment, and requested the technician's inference concerning equipment operation based on that effect. If the technician strayed from the active path during troubleshooting and did not return in a timely manner, the expert presenter provided coaching. The action-result cycle continued until the problem was solved. This procedure was followed for each problem in the problem set. See Table 2 for a sample of three steps drawn from a solution given for the ASP44 problem.
Table 2
Three steps drawn from a solution given for the ASP44 problem. (Problem statement: In debrief, the crew chief reports an ASP44.) Note: P = Precursor, A = Action, R = Result, and I = Interpretation.

<table>
<thead>
<tr>
<th>PARI</th>
<th>Technician Response</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
</tr>
<tr>
<td>P:</td>
<td>Sometimes ASP44 will come down in flight. Also, there was no BIT light and the pilot didn't report any discrepancies.</td>
</tr>
<tr>
<td>A:</td>
<td>Run up aircraft, and see if ASP44 clears.</td>
</tr>
<tr>
<td>R:</td>
<td>ASP44 does not clear.</td>
</tr>
<tr>
<td>I:</td>
<td>Most likely the LRU6 is bad.</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
</tr>
<tr>
<td>P:</td>
<td>LRU6 is probably bad.</td>
</tr>
<tr>
<td>A:</td>
<td>Remove and replace the LRU6.</td>
</tr>
<tr>
<td>R:</td>
<td>ASP 44 latches. No BIT light.</td>
</tr>
<tr>
<td>I:</td>
<td>Something is causing the ASP44 to latch.</td>
</tr>
<tr>
<td><strong>Step 3</strong></td>
<td></td>
</tr>
<tr>
<td>P:</td>
<td>Fault indications go through the LRU3 before being sent to the ASP.</td>
</tr>
<tr>
<td>A:</td>
<td>Remove and replace the LRU3.</td>
</tr>
<tr>
<td>R:</td>
<td>ASP44 does not clear.</td>
</tr>
<tr>
<td>I:</td>
<td>Must have a bad wire somewhere.</td>
</tr>
</tbody>
</table>

**PARI-Stage VII.** During the seventh stage of PARI, the expert problem developers review the problem set to determine the adequacy with which the developed problems assess the cognitive skills and knowledge required for skilled performance. The two expert developers reviewed the problems and ascertained that they were representative of problems encountered in the actual job environment. In addition, they determined that solutions to the problems
required the types of cognitive skills and knowledge associated with skilled performance. In addition to making these judgments, one of the experts rank-ordered the troubleshooting problems in terms of their difficulty (see Table 3). His rank-ordering was based on his observations of the technicians troubleshooting the problems, in addition to his 12 years of experience in the C Shop. The two additional technicians who had worked to solve the developed problems were unable to solve the following three problems without coaching from the expert presenter: ICS Frequency Holes, ASP44, and CMD Safety Switch. Thus, these three problems were considered more than moderately difficult and were dropped from consideration, leaving the RWR-RF Loss problem as the most likely candidate for problem selection.

Table 3  
*Rank ordering of troubleshooting problems from most (1) to least (10) difficult offered by one of the expert problem developers.*

<table>
<thead>
<tr>
<th>Troubleshooting Problem Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ICS Frequency Holes</td>
</tr>
<tr>
<td>2. RWR--RF Loss</td>
</tr>
<tr>
<td>3. ASP 44</td>
</tr>
<tr>
<td>4. CMD Safety Switch to Scope</td>
</tr>
<tr>
<td>5. RWR Wiring--Broken Wire</td>
</tr>
<tr>
<td>6. AAI--Cable at Radar</td>
</tr>
<tr>
<td>7. IFF Seat Switch</td>
</tr>
<tr>
<td>8. KY-58 Relay Panel</td>
</tr>
<tr>
<td>9. 51 &amp; 52 Switches--AAI</td>
</tr>
<tr>
<td>10. Have Quick Word of Day</td>
</tr>
</tbody>
</table>
PARI-Stage VIII. During the eighth stage of PARI, less-skilled technicians are observed as they work to generate solutions for the developed problems. Thus, intermediate and novice technicians were asked to troubleshoot the remaining problems, using the PARI problem solving structure described above. One of the expert problem developers presented the problems he had developed to the technicians. The remaining problems were presented by an expert from the Air Force research team (the expert problem developer was unavailable). Observations of the troubleshooting behavior of these technicians revealed that novice technicians could isolate the fault associated with RWR--RF Loss problem quickly. In addition, they could easily solve the problem if they had previous experience with a specific cable measurement tool, although this tool was only just being introduced to the C Shop and was not in common use. Because the less-skilled technicians could quickly and easily isolate the fault, the RWR--RF Loss problem was deemed inappropriate for selection as the moderately difficult troubleshooting problem. On the other hand, observations of technicians solving the RWR Wiring--Broken Wire problem indicated that this problem revealed performance differences across technicians of different skill levels. Thus, this problem was selected for use as the troubleshooting problem in this study.

The RWR Wiring--Broken Wire Problem

The system important in troubleshooting the RWR Wiring--Broken Wire problem is the Radar Warning Receiver (RWR) system. The RWR is part of the F-15 Tactical Electronic Warfare System (TEWS) and is designed to detect, analyze, and identify threat radar signals. The RWR also controls countermeasure responses to those threats (e.g., the release of chaff).
The RWR System Components. The RWR system consists of eleven components (see Table 4; Gouley, 1992). A functional block diagram of the RWR system components is illustrated in Figure 1. In this section, the general functions of each of the RWR system components are discussed. First, the LRU2 (Line Replaceable Unit) provides the RWR system with its power requirements. Specifically, the LRU2 receives 115 VAC 3 Phase 400 HZ and +28 VDC power from the aircraft and converts it to the DC voltages needed by the RWR. In addition, the LRU2 stores the flight program required for normal operation of the RWR. Next, the LRU3, in conjunction with the low-band antenna, provides the aircraft with coverage in the low-band frequency range. In addition, the LRU3 processes high- and low-band receptions, distributes data between components of the RWR, and interfaces the RWR with other avionics systems. The LRU6, in conjunction with the four high-band antennas (left wing tip, right wing tip, left fin, and right fin), provides the aircraft with omni-directional coverage in the high-band frequency range. The LRU9 provides all displays for the TEWS. It is located in the cockpit and provides the pilot with a view of the threat environment relative to the aircraft. The LRU10 is a control panel which allows the pilot to select on or off for the following systems: RWR, Internal Countermeasure Set (ICS), and Electronic Warfare Warning System (EWWS). The LRU11 is another control panel which allows the pilot to select RWR/ICS combat/ training mode, mode of the TEWS pod, and ICS mode of operation.

Fault Location. The problem statement for the RWR Wiring–Broken Wire is listed in Table 5. This problem is caused by a shorted video cable between the LRU3 and the LRU9. Specifically, the video cable between LRU3-2J8 and LRU9-9J1 is shorted. Center pin 1 on this cable is shorted to (touching) the shield on the LRU3 end, resulting in a loss of data and a blank TEWS display.
Table 4
*RWR System Components.*

<table>
<thead>
<tr>
<th>Component Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRU 2 (Power Supply)</td>
</tr>
<tr>
<td>LRU 3 (Low Band Receiver Processor)</td>
</tr>
<tr>
<td>LRU 6 (High Band Receiver)</td>
</tr>
<tr>
<td>LRU 9 (TEWS Display)</td>
</tr>
<tr>
<td>LRU 10 (TEWS control panel)</td>
</tr>
<tr>
<td>LRU 11 (TEWS immediate action control panel)</td>
</tr>
<tr>
<td>Left Fin Antenna</td>
</tr>
<tr>
<td>Right Fin Antenna</td>
</tr>
<tr>
<td>Left Wing Tip Antenna</td>
</tr>
<tr>
<td>Right Wing Tip Antenna</td>
</tr>
<tr>
<td>Low-Band Antenna</td>
</tr>
</tbody>
</table>

on the LRU9 (see Figure 2). When functioning correctly, this cable transmits video data from the LRU3 to the LRU9. The transmitted data contains information about the type and placement of symbols on the LRU9’s display. However, due to the short between pin 1 and the shield, the data to be transmitted to the LRU9 is lost, and the LRU9 is completely blank.

Based on the problem statement (Table 5), technicians troubleshooting this problem may examine the following areas as possible causes of the problem: LRU9, LRU3, LRU2, or aircraft wiring. Without properly diagnosing the symptoms, the technician may leave the active signal path. For instance, the technician may perceive that the scope is blank because there is no power to the system. Such a perception may lead to an investigation of the circuit.
breaker panel, TEWS control panel, or the LRU2. Conversely, correct identification of symptoms might lead the technician to a "fault code" (available to technicians as part of the Technical Order or T.O.) which specifies a path for isolating a fault given a particular set of symptoms. The fault code appropriate for the RWR Wiring--Broken Wire problem lists the procedure to be followed given the following symptoms: no cross symbol or azimuth dots on TEWS display when power is applied to the RWR. One step in the fault isolation (F.I.) tree for this fault code instructs the technician to check the data cable between the LRU3 and LRU9. Here, it is important that the technician has knowledge
about the correct testing procedure of this cable. The F.I. only specifies to test for continuity across the center conductors from end-to-end for this cable. However, in order to detect the fault, the technician must thoroughly test the cable, which would include checking for continuity from center conductor to shield. Such a test would reveal that there is continuity here, indicating a short.

Table 5
Problem Statement for the RWR Wiring--Broken Wire Troubleshooting Problem.

"In debrief, the pilot reports that the RWR is inoperative, the BIT (built-in test) light is on, and the TEWS display is blank."

![Diagram of cable with pin 1 shorted to shield]

Figure 2. Cross-sectional view of the LRU3 end of the video cable between LRU3-2J8 and LRU9-9J1, in the context of the RWR Wiring--Broken Wire problem. Center pin 1 is shorted to the shield.

Subjects

Subjects were nineteen technicians in the flightline avionics (C Shop) career field of the Air Force. Each subject had been through a technical training school designed to prepare them for their specialty and had received
subsequent on-the-job training. Technicians working in the C Shop are responsible for the identification, isolation, and repair of airborne avionics equipment systems, including the Radar Warning Receiver (RWR) and Identification--Friend or Foe (IFF) subsystems. Technicians were selected so that a range of proficiency levels would be represented. Thus, six 3-levels, eight 5-levels, and five 7-levels participated. All participating technicians but one were male.

Materials and Procedure

Technicians' mental models of an avionics system (i.e., the RWR) were measured using four techniques: a laddering structured interview, concept relatedness ratings, a diagramming structured interview, and think-aloud troubleshooting protocol. Accuracy and time measures were not used for several reasons. First, the performance measure used as the criterion for the evaluation of the mental model measures relied on errors as one performance index. Thus, troubleshooting problem accuracy and the method used for creating the criterion score are not independent, rendering accuracy unusable as a mental model measurement technique. Time measures were considered inappropriate because of the manner in which technicians were asked to troubleshoot the selected problem. Specifically, the troubleshooting session was conducted verbally, with the technician verbally specifying the desired actions and an expert verbally specifying the results of those actions. The time taken to verbalize a particular action is probably much different from the time taken to actually execute that action. In addition, a technician may be able to verbalize an action easily without being able to efficiently execute the same action. Furthermore, time measures do not offer direct information regarding the cognitive state of the measured person. Instead, cognitions are inferred.
Specifically, it is assumed that persons who complete tasks quickly have better, more complete mental models of the involved system. However, it may also be that efficient performance is due to a memorized procedure or simple luck (i.e., the solution was stumbled upon). Thus, time measures were not gathered.

In addition to the four mental model measures, technicians' knowledge was also measured using a retrospective interview, but these data were collected for the purposes of the Air Force research team and will not be discussed here. All mental model measures took place within the context of the RWR Wiring-Broken Wire troubleshooting problem discussed earlier. (The RWR system is the avionics system important for troubleshooting this problem.) In addition to completing all of the mental model measures, each technician worked to verbally troubleshoot the RWR Wiring-Broken Wire problem.

**Laddering Structured Interview.** Upon arrival to the testing session, the technician completed the laddering structured interview. This interview consisted of four steps. The problem statement for the troubleshooting problem was presented (see Table 5) to the technician. The first step of the interview consisted of asking the technician to identify the major system important in troubleshooting this problem. In Step 2 the technician was asked to name the major components of the identified system, in the context of the troubleshooting problem. In the third step of the interview the technician was asked to name all of the major components of the identified system, regardless of problem context. In the fourth and final step the technician was asked to name the major systems with which the identified system interfaced, if any. Throughout these steps an index card with the identified component or system name written on it was prepared and placed before the technician, according to the arrangement specified by the technician. The index cards only served as memory aids for the
technician and were discarded after the identified systems and components had been recorded.

**Relatedness Ratings.** Following the laddering structured interview, the technician completed familiarity ratings (i.e., 1 = highly familiar to 6 = unfamiliar) on the eleven RWR system components (see Table 4). The components were presented in a random order. The technician then completed relatedness ratings on all pairs of the same eleven RWR system components. The technician was told to make all relatedness ratings within the context of the RWR Wiring--Broken Wire troubleshooting problem. The technician was told to rate the items in terms of their functional relatedness. Furthermore, the technician was told that although two items can be functionally related in a number of different ways (e.g., by information flow through the system, by performing similar function), ratings should be made based on the first general impression of functional relatedness of the concepts, within the context of the troubleshooting problem.

Both the familiarity and relatedness ratings were completed on a computer; a HyperCard program was used to collect these data. The technician rated the relatedness of the component pairs by using a mouse to point and click on one of five sections on a bar, with the endpoints labeled "slightly related" and "highly related". If the technician wished to rate the component pairs as unrelated, a button labeled "unrelated" was available. A similar procedure was used to collect the familiarity ratings. The RWR Wiring--Broken Wire troubleshooting problem statement (see Table 5) was available for review at the top of the computer screen throughout the relatedness ratings task. Presentation of pairs was randomized across subjects, and order of items within pairs was counterbalanced.
Diagramming. After completing the ratings, the technician completed a diagramming task, using the component set just rated. Randomly ordered index cards with the name of an RWR component printed on each were given to the technician, as was a set of directional and bi-directional white arrows. The technician was instructed to arrange and connect the components in a manner representing the actual function of the RWR system, in general. The technician specified directionality of relations with the unidirectional and bi-directional arrows. To illustrate the use of the arrows for representing functional relations, the technician was given an example diagram with an accompanying explanation from the domain of automobile engines (see Figure 3).

![Diagram of Piston, Connecting Rod, Crank, and Crank Shaft]

Figure 3. Example functional diagram from the domain of automobile engines.

After the technician completed the functional diagram representing the RWR system at a general level, the RWR Wiring--Broken Wire troubleshooting problem statement (see Table 5) was re-presented. The technician was then given a set of directional and bi-directional yellow arrows and was asked to use these yellow arrows to designate those components and/or connections most important in troubleshooting the RWR Wiring--Broken Wire problem. Finally, the technician was asked to explain, in his own words, both diagrams. These
explanation data were collected to aid the examiner's understanding of the generated diagrams, and no further analyses were conducted on the explanation data.

**Think Aloud.** The technician then proceeded to the troubleshooting/think aloud portion of the experiment. The technician was told that he would be verbally troubleshooting the problem used in each of the previous tasks, the goal being to isolate the fault and repair the equipment. The technician was instructed to think aloud continuously while working to solve the problem, verbally expressing all thoughts. Two practice think-aloud problems were then reviewed with the technician to ensure that the technician understood what was meant by thinking aloud (i.e., what is the result of multiplying 24 by 6, and how many windows are there in your house?). If the technician had difficulties during these practice problems (e.g., did not speak), he was guided to think aloud. After successfully completing these practice think-aloud problems, the verbal troubleshooting session began.

The examiner re-presented the RWR Wiring--Broken Wire problem statement (see Table 5). All technical materials necessary for troubleshooting the problem were available. These materials included the C Shop Job Guide (J.G.) and the T.O. which contains fault isolation trees and schematic diagrams of the RWR system. The technician was instructed that the goal was to isolate the fault and repair the equipment through a series of iterative action-result steps. In addition, the technician was reminded to verbally express all thoughts while working to solve the problem. An expert assisted with this task by "simulating" the equipment for the technician. Specifically, the expert provided the technician with results for all specified actions. The technician began by specifying the first action he would take in troubleshooting the problem (e.g.,
check the Avionics Status Panel or ASP). The expert responded by informing the technician of the action's result (e.g., ASPs 5 and 49 are latched). The technician then specified the next action, the expert gave the corresponding result, and so on. This action-result cycle continued until the problem was solved, the 45-minute time limit expired, or the technician gave up. All of the technician's, the expert's, and the examiner's responses were recorded with an audio tape recorder.

After completing the troubleshooting/think aloud task, the examiner and the technician reviewed the protocol action by action. The technician was asked to provide a retrospective report of: (1) why each action was taken, and (2) the information provided by the corresponding result of each action, in terms of the equipment in question. This retrospective interview was conducted for other purposes and will not be further discussed here.

**Questionnaires.** Following the troubleshooting action review, the technician completed two questionnaires. First, the technician completed a Likert-style questionnaire designed to allow an evaluation of the mental model measurement techniques (see Appendix A). Specifically, the technician used a 6-point scale to rate each of the techniques on the following dimensions: difficulty (difficulty--easy), similarity to actual troubleshooting in the shop (different--similar), range of responses available (restricted--broad), realism relative to actual troubleshooting in the shop (artificial--realistic), and usefulness for measuring system knowledge (useless--useful). Space was available for additional written comments.

The technician was then given a questionnaire designed to allow a comparison of the measurement techniques (see Appendix B). All pairs of tasks (measurement techniques and troubleshooting task) were presented on this
questionnaire. The technician was asked to make pairwise judgments of the tasks by circling the task in the presented pair which best measured knowledge needed for actual troubleshooting of the RWR Wiring--Broken Wiring problem in the shop. Immediately after circling one of the tasks in the presented pair, the technician used an 8-point scale to rate the similarity of the knowledge measured by the circled task to knowledge needed for the actual troubleshooting of a problem in the shop. The endpoints of the scale were labeled "Not at all similar" and "Extremely similar." Upon completion of this questionnaire, the technician was debriefed and excused. Separate from the testing session, a supervisor rating questionnaire was completed for each technician by his respective supervisor (see Appendix C).

DATA ANALYSIS

Overview of Analyses

The data were analyzed to evaluate the various mental model measurement techniques in terms of their abilities to predict differences in troubleshooting performance. The "bottom line" in real-world troubleshooting situations is performance. Thus, comparisons of the results obtained from the measurement techniques with the results from the troubleshooting task should provide a pragmatic means of assessing the validity of the techniques. This approach is not straightforward however, because the results of each of the techniques offer only information regarding the content of technicians' knowledge, whereas performance is generally indicated in terms of reference to some ideal or perfect score. Therefore, what is needed is an assessment of the technicians' knowledge, using the mental model measure. Such an
assessment requires an ideal or "gold standard." In general, the gold standard for each measure was based on that measure applied to a group of four high performers (i.e., technicians who had the highest verbal troubleshooting scores). The group's combined results comprised the gold standard for that measure. In cases in which there was little agreement across high performers, the group was limited to those who agreed. The performance criterion was based on the outcome of the subjects' performance on the verbal troubleshooting task. In the following section, the method used to generate a troubleshooting performance score for each technician (i.e., the criterion value) is described. This approach to scoring is the current assessment method in the domain of avionics troubleshooting (Pokorny & Gott, 1994).

**Troubleshooting Performance**

Two subject matter experts independently scored the technicians' troubleshooting action protocols. These experts had participated in the troubleshooting problem development stage. In addition, one of the experts (Expert B) assisted the examiner in problem presentation when the technicians worked to troubleshoot the problem. The protocols given to the experts for scoring included the actions executed by the individual technicians, along with their corresponding results. The experts were instructed to score the protocols based only on the actions which were listed. They were told to use a 100-point scale (100 = perfect). In addition, the experts were asked to rank-order the protocols from best (1) to worst (19). Figures 4a and 4b represent the troubleshooting performance score distributions for Experts A and B, respectively.
Figure 4. Troubleshooting performance score distributions for Expert A (Figure 4a) and Expert B (Figure 4b).

An examination of the Figure 4 reveals that the two experts used the scale differently in scoring the troubleshooting protocols. Expert A's scores ranged from 0 to 100, whereas Expert B's scores ranged from 65 to 100. However, the scores awarded by the two experts were significantly correlated\(^1\) \( r (17) = .883, \ p < .0001\), as were the rank-orderings given by the experts, Spearman's \( r (17) = .847, \ p < .0005\). Thus, a single performance score was created for each technician by averaging the troubleshooting scores given by the two experts (see Figure 5). This score was then used as the performance score in all subsequent analyses. Four high performers emerged from this analysis who had performance scores one standard deviation greater than the mean (i.e., all scores 95 or above). These four technicians were used to create the gold standard for comparison in the remaining analyses.

\(^1\) All correlations reported in this paper are Pearson product moment correlations unless otherwise specified. In addition, because correlations between performance and knowledge were expected in particular directions, they were tested for significance using one-tailed probabilities.
Figure 5. Frequency distribution of the average troubleshooting performance score awarded to each technician.

RESULTS AND DISCUSSION

Laddering Structured Interview

To determine the level of inter-subject agreement among the four high performers, the proportion of shared items across lists for each pair of high performers was calculated for each step in the interview (see Table 6). That is, the ratio of shared items to the total number of different items listed was calculated for each step for all pairs of high performers. The resulting proportions indicate that the high performers agreed on the important components or systems for each step, particularly Steps 1 and 2. All high performers listed the RWR as the system important in troubleshooting the RWR Wiring--Broken Wire problem (Step 1). In addition, for Step 2 each of the high performers named at least three of the same components (i.e., the LRU2, LRU3, and LRU9) as important components in troubleshooting this problem.

A gold standard component/system list was created from the lists of these four high performers for each of the four steps (see Table 7). Items named by at
least one of the high performers were included in the list. Knowledge indices for the 15 remaining technicians were then calculated in terms of the proportion of items shared with the gold standard list associated with each step of the interview. The resulting knowledge indices for each step were correlated with the troubleshooting performance score, excepting Step 1 (see Table 8). At Step 1, all technicians named the RWR as the system important in troubleshooting.

Table 6
The proportion of shared list items for pairs of high performers for each interview step.

<table>
<thead>
<tr>
<th>Technician Number</th>
<th>6</th>
<th>8</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>5</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Step 2</td>
<td>1.0</td>
<td>.75</td>
<td>.50</td>
</tr>
<tr>
<td>Step 3</td>
<td>.67</td>
<td>.63</td>
<td>.71</td>
</tr>
<tr>
<td>Step 4</td>
<td>.50</td>
<td>.67</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>6</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td></td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td>.67</td>
<td>.50</td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td>.63</td>
<td>.60</td>
</tr>
<tr>
<td>Step 4</td>
<td></td>
<td>.40</td>
<td>.50</td>
</tr>
<tr>
<td><strong>8</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td>.50</td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td>.67</td>
</tr>
<tr>
<td>Step 4</td>
<td></td>
<td></td>
<td>.67</td>
</tr>
</tbody>
</table>
Table 7
Gold standard lists for each interview step. Note: The Step 3 list includes items named by individual technicians as major components of the identified system, regardless of problem context, in addition to those items named by that technician in Step 2.

<table>
<thead>
<tr>
<th>Laddering Interview Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
</tr>
<tr>
<td>RWR</td>
</tr>
<tr>
<td>LRU3</td>
</tr>
<tr>
<td>LRU6</td>
</tr>
<tr>
<td>LRU9</td>
</tr>
<tr>
<td>LRU10</td>
</tr>
<tr>
<td>Aircraft Wiring</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

the problem, indicating that all sampled technicians correctly interpreted the presented symptoms at a coarse level. Correlations conducted on the remaining three steps indicated that naming components important for troubleshooting the problem (Step 2) was predictive of troubleshooting performance, $r(13) = .542, p < .025$. This positive relationship indicates that good troubleshooters agreed with the high performers on the components important for troubleshooting the problem more than did the poor troubleshooters. The data resulting from Steps 3 and 4 (i.e., name all
Table 8
Correlations between troubleshooting performance and knowledge indices: 1) the proportion of shared items with the gold standard list and 2) errors of commission for Steps 2-4 of the laddering interview.

<table>
<thead>
<tr>
<th></th>
<th>Proportion of Shared Items</th>
<th>Errors of Commission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 2</td>
<td>.542</td>
<td>.505</td>
</tr>
<tr>
<td>Step 3</td>
<td>-.128</td>
<td>-.326</td>
</tr>
<tr>
<td>Step 4</td>
<td>-.200</td>
<td>-.296</td>
</tr>
</tbody>
</table>

components regardless of context, and name interfacing systems) were not predictive of troubleshooting performance (see Table 8).

Note, the multistage Bonferroni procedure (Larzelere & Mulaik, 1977) was used to control for inflation of the Type I error rate within tests conducted on the Laddering interview steps measure. Alpha family-wise ($\alpha_{FW}$) was set at .10. Only correlations tested at the first stage of the procedure were significant, with alpha test-wise ($\alpha_{TW}$) = .03. Likewise, in the analyses that follow each mental model measure was treated as an independent measure, and this same procedure for protecting against inflation of the Type I error rate was followed for each measure. Only correlations tested at the first stage of the procedure for each measure were significant, thus only one $\alpha_{TW}$ value is reported.

In addition to examining the proportion of shared components/systems, the number of items listed by technicians, but not by the high performers, was calculated for each interview step for each technician. These errors of commission made up a second set of knowledge indices which were then correlated with troubleshooting performance. Again, data from Step 1 were not included in these correlations because all technicians named the RWR as the
important system in troubleshooting the problem, and no errors of commission were made. Of the remaining three steps, only errors of commission occurring during Step 2 of the interview were predictive of troubleshooting performance, \( r (13) = .505, p < .05 \). This marginally positive correlation indicates that technicians who identified extra RWR system components important in troubleshooting the problem (not included in the gold standard list) were better troubleshooters than were technicians who did not. The correlations resulting from the remaining steps were not significant (see Table 8, \( \alpha_{TW} = .03 \)).

In general, the second step of the laddering structured interview is significantly related to troubleshooting performance. Those technicians who listed more components that were shared with the gold standard list performed the troubleshooting task better than those who listed fewer gold standard components. Unexpectedly, those technicians who also listed more components not on the gold standard list were also better troubleshooters than those who listed fewer. This latter result, although unexpected, corroborates earlier findings in this area. Specifically, Cooke and Rowe (1993) found that as students gained troubleshooting experience, they tended to execute a greater number of actions (and even more than high performers), although they did not seem to know when the actions should be applied. In a similar way, the best troubleshooters in this study believed that many components were relevant for troubleshooting the problem, including those which are actually relevant. Perhaps, early stages of the development of expertise can be characterized by a familiarity with many components and procedures, whereas the mapping of those components and procedures to a particular troubleshooting situation is a hallmark of later stages of expertise.
Finally, it is interesting that the laddering technique was predictive of troubleshooting performance only in the context of the troubleshooting problem. Lists of the general system components or interfacing components were not predictive. Interestingly, Gitomer (1984) also used a laddering structured interview in this domain, however he did not restrict the interview to a particular problem context. Instead, subjects were told to think of a specific LRU and to iteratively break this LRU down into its components. This lack of context may explain why Gitomer did not observe a difference between skilled and less-skilled airmen (as defined by supervisor ratings) in the laddering technique.

Relatedness Ratings

Technicians' familiarity ratings indicated that, on average, they were familiar with the RWR system components ($M = 1.9$, $SD = .919$, collapsing across components and technicians). However, two 3-level technicians each rated two components as unfamiliar (rating = 6). First, they both rated the LRU11 as unfamiliar. This component is a control panel with which technicians do not regularly interact. Perhaps these two technicians had not yet come into contact with the LRU11 in the course of troubleshooting. Furthermore, it is not an important component for troubleshooting the RWR--RF Loss problem. The two technicians also rated different antennas as unfamiliar. One technician rated the low-band antenna as unfamiliar, and one rated the right fin antenna as unfamiliar. These ratings are unusual, particularly the rating given for the right fin antenna. The right fin antenna is one of the four high-band antennas. The technician rating the right fin antenna as unfamiliar rated the remaining three high-band antennas as familiar. These antennas perform the same function in different areas of the aircraft. Being familiar with one of the high-band antennas implies familiarity with the remaining three high-band antennas. Perhaps these
ratings could be attributed to input error. Regardless, on average, the technicians were very familiar with the RWR system components.

Correlations of relatedness ratings for each pair of the four high performers were computed to determine degree of inter-subject agreement. These correlations are presented in Table 9. Note that the correlations are all high and statistically reliable ($p < .05$ with 54 degrees of freedom).

Table 9

<table>
<thead>
<tr>
<th>Technician Number</th>
<th>6</th>
<th>8</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>.757</td>
<td>.769</td>
<td>.889</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>.662</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td>.899</td>
</tr>
</tbody>
</table>

In order to generate a graphical summary of the ratings, the data were submitted to the Pathfinder network scaling procedure, a descriptive multivariate statistical technique that represents pairwise proximity in a network form (Schvaneveldt, 1990). In the networks, concepts (or components in this case) are represented as nodes, and relations (functional relations in this case) are represented as links between nodes. This network representation not only summarizes the data but has also been shown to convey information about conceptual relatedness not seen in the ratings themselves (Cooke, 1992; Cooke, Durso, & Schvaneveldt, 1986). For more detail on Pathfinder see Schvaneveldt (1990). The C statistic (Goldsmith & Davenport, 1990), a measure of shared links for matching nodes across two different networks, was
calculated between each of the four high performers (see Table 10). This measure ranges from 0 (low similarity) to 1 (high similarity) and can be viewed as a measure of association between two networks. The C values in Table 10 indicate that technician #8 shares fewer links with the other three high performers than the other three share with each other. Therefore the gold standard was computed in several ways: (1) averaging ratings of all four high performers, (2) using ratings of technician #8 only, and (3) averaging the ratings of all high performers, excluding #8. The resulting Pathfinder network based on the data from the second gold standard is presented in Figure 6a.

Table 10

<table>
<thead>
<tr>
<th>Technician Number</th>
<th>6</th>
<th>8</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>.887</td>
<td>.418</td>
<td>.868</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>.423</td>
<td>.759</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td>.404</td>
</tr>
</tbody>
</table>

The knowledge index for each of the 15 technicians was based on the C value between the individual technician's network and each of the three gold standards described above. For comparison purposes Figure 6b presents the Pathfinder network based on the mean relatedness ratings of the eight lowest performers (i.e., troubleshooting score less than 50). Correlations were then computed between knowledge indices and troubleshooting scores to determine which, if any, of these three knowledge indices predicted performance. Correlations were respectively, .399, .527, and .425 for the first, second, and
third gold standards described above. With 13 degrees of freedom, the only predictive gold standard was the one based on technician #8 alone ($\alpha_{TW} = .03$). A partial correlation ($r (12) = .405, p = .06$), although only marginally significant, suggested that technician #8 was predictive of performance even when the variance accounted for by the remaining three high performers (i.e., the third gold standard) was partialled out. On the other hand, the correlation between performance and the gold standard based on all four high performers (i.e., first gold standard) dropped from .399 to -.034 when the variance due to technician #8 was partialled out.

![Figure 6a](image1)

![Figure 6b](image2)

**Figure 6.** Pathfinder networks ($r = \infty, q = n-1$) based on: a) the data from technician #8 - the second gold standard and b) the mean relatedness ratings given by the eight lowest performers (score < 50). Note: 1 = highly related, 6 = unrelated.

These results indicate that relatedness ratings, coupled with the Pathfinder network analysis procedure, are only marginally predictive of troubleshooting performance when all high performers are used as the gold standard. More
specifically, however, these results suggest that one of the high performers, technician #8, generated relatedness ratings that were significantly predictive of performance, independent of the remaining high performers. The C values for the high performers, in combination with the marginally significant partial correlation, indicate that there is little overlap between the network of technician #8 and those of the other three high performers. Interestingly, there are some other differences between technician #8 and the other high performers. That is, #8 is a 5-level, whereas the others are all 7-levels. Also, #8 has spent only four years in the C shop, whereas the others have spent between 6 and 8 years in the C shop. However, #8 performed as well as the other high performers in the troubleshooting task, receiving a score of 95 compared to the other three scores of 95, 97, and 100. Possible explanations for this intriguing pattern of results are discussed in the General Discussion section.

Diagramming

The following analyses are based on the general diagrams of the RWR system. The diagrams that were specific to the RWR Wiring--Broken Wire problem were not informative because the majority of the technicians deemed that only three to five system components were relevant to this problem. Each of the 19 technician's system diagrams was converted to an 11 by 11 asymmetric matrix, with ones representing the presence of a connection between components and zeros indicating no connection. To determine the level of diagram similarity among the four high performers the proportion of shared connections for each pair of diagrams was computed. This proportion was based on the number of links shared by the two diagrams divided by the total number of links in the union of the two diagrams. These proportions are presented in Table 11. In general, pairs of high performers shared about half of
the links present in the two diagrams. However, closer inspection of Table 11 indicates that once again, the diagram of technician #8 shared the least with those of the other high performers (mean proportion of .42 for technician #8 compared to .68 for other pairs of high performers). For this reason, three gold standard diagrams were created parallel to the three gold standards for relatedness ratings: (1) a diagram based on all four high performers, (2) technician #8's diagram, and (3) a diagram based on all high performers excluding #8. The matrix representing the group diagrams (i.e., gold standards one and three) consisted of ones, indicating that the connection existed in at least one diagram, and zeros otherwise. Again these matrices were asymmetric.

Table 11
Proportion of shared connections for pairs of high performers' diagrams.

<table>
<thead>
<tr>
<th>Technician Number</th>
<th>6</th>
<th>8</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>.55</td>
<td>.48</td>
<td>.85</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>.39</td>
<td>.65</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td>.40</td>
</tr>
</tbody>
</table>

Three diagramming knowledge indices were generated for each of the remaining 15 technicians by subtracting the technician's matrices from each of the three gold standard matrices and summing the absolute values of the differences. These indices should be zero if there is complete agreement with the gold standards. The correlations between these knowledge indices and trouble shooting performance were -.464, -.440, and .089 for gold standard
one, two, and three, respectively. The first two correlations are marginally
significant ($p = .041$ and .051, respectively, $\alpha_{TW} = .03$) with 13 degrees of
freedom. Note that a negative correlation is expected given that large
knowledge indices indicate large diagram differences. The low correlation
between gold standard three and performance, as well as a nonsignificant
correlation of -.181 when technician # 8's diagram is partialled out of the first
gold standard, suggests that the predictability of the high performers in this
technique can be attributed to technician #8. In fact, when the three other high
performers are partialled out of the correlation between technician #8 and
performance, the resulting correlation is high and significant ($r (12) = -.538, p=
.02$). Thus, as for relatedness ratings, technician #8 is more predictive of
technician performance than the other three high performers. The diagram
completed by technician # 8 is presented in Figure 7.

Interestingly, using #8 as the gold standard, errors of omission ($r (13) =
-.508, p =.026$) are better predictors of performance than errors of commission
($r (13) = -.272, p=.167$). This result indicates that those technicians whose
diagrams did not include connections seen in Technician #8's diagram tended
to exhibit poorer troubleshooting performance. It is interesting that errors of
omission should be predictive as opposed to commission, given the laddering
task finding in which errors of commission were more predictive of performance.
However, these two results may not be disparate. Specifically, the fact that
errors of omission in the diagramming task were predictive suggests that seeing
"extra" system relations is not as problematic for troubleshooting as is failing to
see one or more critical relations. Similarly, the fact that errors of commission in
the laddering task were predictive suggests that the best troubleshooters
believed that many components were relevant for troubleshooting the problem,
including those which are actually relevant. Thus, better troubleshooters appear to have a wide range of information available, a subset of which is considered critical by experts. These results also indicate that the laddering and diagramming tasks may be accessing different aspects of mental model knowledge.

Figure 7. RWR system diagram created by technician #8.

In general, the diagramming technique, at least in the context of the overall system, predicted troubleshooting performance well. Most of this predictability can be attributed to technician #8, the same technician whose relatedness ratings predicted performance. In terms of the diagramming technique, technician #8 differed from the other high performers in two major ways: (1) #8 used only bi-directional arrows, and (2) based on technicians' diagram explanations, #8 attempted to represent both information flow and power flow in his diagram, whereas the other high performers represented only information flow in their diagrams.
Think Aloud

A coding scheme for technicians' verbalizations, as well as their actions in the think-aloud-while-troubleshooting technique was developed. The purpose of the coding scheme was to classify verbalizations and actions into discrete meaningful units that could be represented as nodes in a Pathfinder network. The main groups of verbalizations included: (1) action interpretation/explanation, (2) result interpretation, (3) component elimination, (4) elimination justification, (5) plan/prepare for test/check, and (6) Technical Order (T.O.) search/interpretation. The most abstract level of categorization was used for each category except for the action interpretation/explanation category. This category was broken down into sub-units because the type of action interpretation/explanation seemed important in distinguishing skill levels. For example, if a technician checked the fuses on the LRU2 (a power check) and verbalized an explanation other than a power explanation, this information would likely distinguish different skill levels and should be captured. The resulting coding scheme consisted of 22 verbalization units/nodes. Each technician's verbal protocol was coded by two raters. The inter-rater reliability achieved on the recoding of the action interpretation/explanation category was acceptable, with 92.6% agreement on 149 coded explanation/interpretation verbalizations. The raters discussed the 11 verbalizations on which they disagreed, and a compromise was made.

Also included in the coding scheme were meaningful troubleshooting actions that technicians executed. These were included in order to provide context for the verbalizations. Meaningful was defined as actions indicative of skill in troubleshooting the presented problem. For example, checking the LRU9 ETI meter (a power indicator) is indicative of troubleshooting skill for the
RWR Wiring--Broken Wire problem because power to the LRU9 may be the cause of the problem. On the other hand, checking the LRU6 ETI meter is not because problems with the LRU6 are not indicated by the problem statement. The main groups of actions included: (1) debriefing questions, (2) equipment checks, (3) continuity tests, and (4) swaps. The most abstract level of categorization that indicated troubleshooting skillfulness was used. Using this decision rule, an action unit was associated with either poor or good troubleshooting actions. The resulting coding scheme consisted of 75 action units/nodes. Two raters coded 5 of the action protocols together. They then coded the remaining 14 action protocols and achieved an acceptable level of reliability of 98.1% with 267 coded actions. The raters discussed the five actions on which they disagreed, and a compromise was made. The entire coding scheme (verbalizations and actions) included 97 events/nodes (22 verbalizations + 75 actions) and can be seen in Appendix D.

Transition probabilities for all event pairs (verbalizations and actions) were calculated for individual subjects by dividing the frequency with which specific event transitions (e.g., T.O. search/interpretation followed by continuity check between LRU2 and LRU3) occurred by the frequency with which the first event in the sequence occurred. For example, if T.O. search/interpretation occurred twice and was followed by continuity check between LRU2 and LRU3 on one of those occasions, then the transition probability would be 0.5. Note that these are first-order transitions only. Higher-order transitions were not used because the immediate transitions were considered to be the most meaningful for this task. Also note that event transitions convey order information, and thus each event pair can be associated with two distinct transition probabilities.
Transition probabilities were calculated for each of the four high performers. Agreement among these technicians was assessed by correlating these probabilities for all pairs of high performers. These correlations are presented in Table 12. Note that all inter-technician correlations are low and are not statistically significant. However, because there is no single technician who seems different from the rest, the transition frequencies of all four technicians were combined as the gold standard. The gold standard transition matrix was then submitted to the Pathfinder network scaling algorithm, as were transition matrices of each of the individual technicians (Schvaneveldt, 1990). Similarity measures (i.e., C values as used for relatedness ratings) among pairs of networks for high performers corroborated the correlation results, with a mean inter-technician C of .06. This low agreement suggests that high performers conveyed very different thoughts in their verbalizations.

Table 12  
Correlations of transition probability matrices for pairs of high performers.

<table>
<thead>
<tr>
<th>Technician Number</th>
<th>6</th>
<th>8</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>.05</td>
<td>.15</td>
<td>.06</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>.06</td>
<td>.01</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td>.03</td>
</tr>
</tbody>
</table>

C values between the gold standard Pathfinder network and each individual technician served as a knowledge index. This index was not predictive of troubleshooting performance ($t (13) = -.026$). Note that this is not a matter of idiosyncrasies among the high performers in that gold standards
based on any one of the four are not significantly predictive of performance (correlations of performance and knowledge indices based on individuals are -.325, -.005, -.133, and -.139). The highest observed correlation (between technician #5 and performance, r = -.325) is in the opposite direction, indicating that the best troubleshooters verbalized in a manner least like that of technician #5.

An additional knowledge index was derived from a correlation of event frequencies (i.e., the frequency with which each discrete verbalization or action occurred) associated with an individual's protocol and event frequencies associated with the gold standard. Thus, this measure should be high to the extent that the technician exhibited the same verbalizations and performed the same actions as the high-performers, the same number of times. It should overlap with the Pathfinder network similarity measure in that they both take shared events into account. However, these two measures are divergent in that the Pathfinder measure includes information on event sequences, whereas the action frequency measure does not. Conversely, the frequency measure includes frequency of individual events, whereas Pathfinder does not. This measure (r (13) = -.394, p = .07) was slightly more predictive than the network similarity measure, though not significantly so. However, this correlation is negative, indicating that technicians with verbalization/action frequencies similar to that of the gold standard had lower troubleshooting scores.

In summary, the think aloud technique resulted in low levels of agreement among the four high performers and was not predictive of troubleshooting performance. The marginally significant negative correlation between the frequency knowledge index and performance suggests that the better technicians say different things at different rates than high performers.
Specifically, high performers tended to say less than the other technicians. However, of the 15 technicians, those who performed well tended to read the Technical Orders and thus generate more verbal statements. Together these results seem to indicate that the think aloud technique does not assess knowledge that is critical for performance. Technicians may be unaware of much of the knowledge that underlies their troubleshooting performance and thus, when asked to think aloud, they verbalize thoughts that are independent of task performance.

Questionnaires

Each of the questions on the Likert-scale questionnaire was analyzed using a repeated measures ANOVA with the four techniques (i.e., laddering, ratings, diagramming, and think aloud) making up the four levels of the independent variable. Two of the five questions resulted in no significant technique effect. These questions had to do with similarity to actual troubleshooting and usefulness as a measure of system knowledge. The other three questions resulted in a significant technique effect. For the task difficulty question, there was a significant effect of technique ($F(3, 54) = 3.5, p = .02$), that can be attributed to the fact that the laddering technique was rated as significantly easier than each of the other three techniques. Respective means for laddering, ratings, diagramming, and think aloud were 5.3, 4.2, 4.3, and 4.2. For the question about restriction in range of responses, there was also a significant technique effect ($F(3, 54) = 6.5, p = .003$). The ratings technique was viewed as significantly more restrictive than laddering and think aloud techniques. Respective means for laddering, ratings, diagramming, and think aloud were 4.5, 3.6, 4.4, and 5.0. For the task artificiality question there was a significant technique effect ($F(3, 54) = 3.42, p = .02$), with the think aloud and
laddering techniques receiving greater realism ratings than the other two tasks. Respective means for laddering, ratings, diagramming, and think aloud were 4.0, 3.4, 3.4, and 4.6. Also, the 19 technicians, on average, viewed mental model knowledge or system knowledge as extremely important (mean rating of 4.3 on a 5-point scale, SD = .65) for troubleshooting the RWR Wiring--Broken Wire problem.

On the second questionnaire technicians circled the technique in each pair of techniques which they felt best measured the knowledge necessary for troubleshooting. Percent responses indicated that technicians viewed the diagramming (39%) and think aloud (33%) techniques as better in this regard than the laddering (12%) and rating (16%) techniques.

Supervisor ratings were to have been collected for each of the 19 technicians. However, ratings for technicians from one of the participating fighter squadrons were unavailable. Six technicians are members of this squadron. Thus, ratings were collected for only 13 of the technicians. Due to the missing data, supervisor ratings were not analyzed.

**GENERAL DISCUSSION**

**Summary and Recommendations for Measuring Mental Models**

The purpose of this project was to evaluate the validities of several mental model measurement techniques. Air Force technicians with varying degrees of expertise completed four mental model measures. These measures were conducted in the context of a specific troubleshooting problem. Care was taken in the selection of this problem to insure that it was (1) representative of the domain (i.e., C shop avionics), and (2) that it was of intermediate difficulty, presumably making the invocation of a system mental model necessary for
successful troubleshooting. In addition to completing the mental model measures, each technician worked to verbally troubleshoot the problem. Comparing the results of the mental model measures to troubleshooting performance should provide a pragmatic means of assessing the validity of the measures. Of the four techniques tested, all but the think-aloud technique were predictive of troubleshooting performance. Whether the think-aloud gold standard was constructed from individual high performers or groupings of these high performers, predictability remained low. On the other hand, the laddering (Step 2), relatedness ratings, and diagramming techniques were all predictive of troubleshooting performance. (Table 13 lists the correlations between these techniques and troubleshooting performance.) Also, when technician #8 was used as the gold standard in the ratings and diagramming techniques, predictability was optimal. An explanation for this result is discussed below.

Table 13
Correlations of troubleshooting performance (TS Score), laddering technique (Step 2), ratings technique, and diagramming technique.

<table>
<thead>
<tr>
<th>Measurement Technique</th>
<th>Laddering</th>
<th>Ratings</th>
<th>Diagram</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS Score</td>
<td>.542</td>
<td>.527</td>
<td>-.440</td>
</tr>
<tr>
<td>Laddering</td>
<td></td>
<td>.288</td>
<td>-.477</td>
</tr>
<tr>
<td>Ratings</td>
<td></td>
<td></td>
<td>-.274</td>
</tr>
</tbody>
</table>

Partial correlations indicated that two of the techniques were each independently predictive of troubleshooting performance. Specifically, the relatedness ratings technique was predictive of performance independent of both the laddering technique ($t (12) = .461, p = .042$) and the diagramming
technique ($r (12) = .471, p = .039$). Similarly, the laddering technique was predictive of performance independent of the ratings technique ($r (12) = .480, p = .035$) and to a lesser extent, the diagramming technique ($r (12) = .421, p = .057$). Interestingly, the diagramming technique was not predictive independent of the relatedness ratings ($r (12) = -.362, p = .09$) or the laddering ($r (12) = -.246, p = .18$) techniques. Thus, the ratings technique and the laddering technique each accessed mental model information which is independently predictive of performance, whereas the diagramming technique did not.

Based on the results of this research, several recommendations regarding the measurement of mental model knowledge can be made. First, the relatedness ratings and laddering techniques each independently predicted troubleshooting performance, indicating that these two methods are valid means of assessing mental model knowledge. On the other hand, neither the diagramming technique or the think-aloud technique was independently predictive of performance. Thus, researchers should consider using the relatedness ratings and laddering techniques in work requiring the measurement of mental models. Interestingly, this recommendation is counter to the technicians' ratings which indicated that they believed the diagramming and think-aloud techniques were the best measures of mental model knowledge.

Further recommendations can be made based on the subjective ratings given by technicians, as well as the procedural features of these two techniques. Specifically, the laddering technique should be used in cases in which both ratings and laddering cannot be used. Technicians thought that the ratings technique was the most restrictive in terms of response freedom, and that it lacked realism. Procedurally, relatedness ratings are also restrictive in
the sense that presenting all pairs of a concept set quickly leads to an unmanageable number of pairs as the number of concepts increases, making the ratings technique nearly impossible to use with large concept sets. Even with a smaller number of concepts (around 20), the pairwise ratings task seems quite long and tedious to subjects. On the other hand, the laddering technique appears to be easier to implement. The laddering technique requires less background knowledge on the part of the researcher than does the ratings technique, especially when ratings are followed by Pathfinder analyses. In addition, technicians rated the laddering technique as realistic and easy to complete. However, there is a tradeoff to be considered. The ratings technique, while relatively more difficult to implement and analyze than the laddering technique, provides a graphical representation of knowledge that is much richer than the list produced using the laddering technique.

The results obtained from this work are particularly promising when aspects of the study are considered. First, the sample size was minimal. The data were gathered in a "real-world" setting, increasing the ecological validity of the conclusions. However, subjects are not typically available in large numbers in such settings, particularly when they must leave their daily work to participate. Only nineteen subjects participated in this study. Furthermore, data from four of these subjects were extracted for use in construction of the gold standard, effectively restricting the range of data on which the correlations were based. Thus, the correlations reported here may be underestimated. Finally, the purpose of this work must be emphasized: the interest was an evaluation of the validity of several mental model measures, with the criterion being troubleshooting performance. The observed correlations between mental model measure performance and troubleshooting performance varied across
the evaluated techniques, indicating that not all mental model measurement
techniques are created equal. Some measures are predictive of performance,
whereas others are not. Measures predictive of performance should be
pursued in future work. In other words, given a selection of mental model
measures, it would pay to use those measures that produce output that seems
most relevant to performance.

What remains to be done is an in-depth examination of the information
provided by the various methods for measuring mental models. A
characterization of the types of information provided by the different methods for
measuring mental model knowledge would assist researchers in choosing the
most appropriate method or methods to meet their research needs. For
example, the laddering interview task appears to access knowledge of system
components, whereas the ratings task appears to measure knowledge of
relationships among system components. Researchers may find these types of
information differentially useful, depending on their research needs.
Conducting an in-depth analysis of the types of information provided by different
mental model measures would allow researchers to select the measure(s) best
suited to their research needs. Furthermore, such a characterization would
serve to strengthen the network of evidence (or nomological network, Cronbach
& Meehl, 1955) surrounding the mental model construct. For example, different
measures, though equally predictive of performance, may tap different aspects
of mental model knowledge. An understanding of the types of information
provided by the different measures would facilitate the integration of research
results, increasing understanding of the mental model construct in general.
Findings on Measuring Mental Models

Defining the Gold Standard. This research revealed several interesting observations relevant to mental model measurement and expertise in general. The first, and perhaps most important, observation dealt with defining the "gold standard." A single technician, #8, contributed to most or all of the predictability associated with the high performers in the ratings and diagramming techniques. As mentioned earlier, technician #8 differed from the other high performers on two dimensions: years of experience and level obtained in the Air Force classification system. Technician #8 had fewer years of experience and was at a lower level in the Air Force classification system. This technician may be a very good "intermediate" level technician rather than an "expert" level technician. Perhaps technician #8 was a better predictor than the other high performers, because the latter were too far removed from the entry-level technicians in terms of expertise, whereas #8 was not. Thus, although technician #8 and the other three high performers each performed well on the troubleshooting problem, their performance on the knowledge measures was different. Perhaps technician #8's knowledge was most like that of a good intermediate-level technician. Using experts as the gold standard assumes a linear relationship between the development of expert performance levels and the development of knowledge. Perhaps major qualitative changes in knowledge occur as expertise develops, making expert knowledge a poor predictor of novice performance.

Such qualitative changes in knowledge are central to the perspective taken by phase learning theorists (Shuell, 1990). Phase theorists assert that learning a complex body of knowledge involves a series of phases, during which the learning process is fundamentally different. Furthermore, although it
is typically assumed that the phases are organized in a linear manner, a non-linear organization is possible. Thus, changes in knowledge may not be adequately represented as a simple monotonic increase in similarity to some ideal knowledge representation (Acton, Johnson, & Goldsmith, in press). Instead, the knowledge organization of a subject at one expertise level may be qualitatively different from that of another subject at a different expertise level, making unique "gold standards" or ideals necessary for each phase of knowledge development.

An analysis conducted on unpublished data provided by Cooke (1994b) offers support for this point. Cooke collected pairwise similarity ratings on a set of cognitive psychology concepts from undergraduate students enrolled in a cognitive psychology class. Two sets of ratings were collected: one set at the beginning of the semester and one set at the close of the semester. The course instructor also completed the ratings. In order to evaluate the predictive power of different gold standards or ideals, two gold standards were evaluated: (1) the ratings given by the instructor and (2) an aggregate high-performer rating set created by averaging the second set of ratings given by the top 5 students in the class (as determined by final grades). Separate analyses were conducted on the first and second set of ratings given by the remaining students. The second set of ratings given by individual students were correlated with both the instructor’s ratings and the average high-performer ratings. These values were then correlated with final grades to determine if the gold standards differed in their respective abilities to predict performance. The resulting correlations indicated that both gold standards were predictive of class performance, $r_{62} = .486$, $p < .0001$, and $r = .645$, $p < .0001$ for the instructor and the high performer gold standards, respectively.
Partial correlations were then conducted to determine if the gold standards were independently predictive of performance. The resulting partial correlations indicated that the high-performer gold standard was predictive of performance independent of the instructor gold standard, $r (61) = .4859, p < .01$. However, the instructor gold standard was not predictive of performance independent of the high-performer gold standard, $r (61) = .058$. Thus, students who gave ratings similar to those given by the high performers were more likely to succeed in the class. Giving ratings similar to those of the instructor was not independently related to class performance. Interestingly, a similar analysis conducted on the students' first set of ratings indicated a similar trend. Specifically, students whose first set of ratings were like the second set of ratings given by the high performers tended to perform well in the class, $r (62) = .266, p = .04$. On the other hand, there was no relationship between having ratings similar to those given by the instructor and class performance, $r (62) = .152, p = .24$. Taken together, these results indicate that an ideal based upon high-performing students offers more predictive power than an ideal based upon the expert instructor.

Similar results were obtained in a study conducted by Acton, Johnson, and Goldsmith (in press) who examined different structural knowledge referents or gold standards in terms of their abilities to predict exam performance. These researchers collected pairwise similarity ratings on a set of programming concepts from students enrolled in one of three BASIC programming classes. Ratings were also collected from the course instructors and from a group of non-instructor experts. Four gold standard structures were evaluated, including: (1) the course instructor's ratings, (2) individual experts' ratings, (3) an average of the ratings given by the non-instructor experts, and (4) averages of the ratings
given by the six students receiving the highest marks in each of two programming classes.

The results indicated, among other things, that the gold standard based on the six best students in the class was a slightly better predictor of exam scores for that class ($r (26) = .35, p < .10$ and $r (31) = .57, p < .05$ for Classes 1 and 2, respectively) than the gold standard based upon the course instructor ($r (26) = .33, p < .10$ and $r (57) = .55, p < .05$ for Classes 1 and 2, respectively).

Furthermore, when the gold standards based upon the students in Classes 1 and 2 were used to predict performance in a third, more advanced class, the resulting correlations were larger ($r (8) = .46$ and $r (8) = .45$, respectively) than when the course instructor was used to predict performance ($r (8) = .35$), although none of these correlations are significant.

Furthermore, an examination of the correlations resulting from the gold standards based on the non-course instructor experts indicates a similar trend for two of the three classes. That is, the student-based gold standard predicted in a manner similar to or slightly higher than the gold standards based on the non-instructor experts. However, in one class the non-instructor expert gold standards were more predictive of performance. In general, these results indicate that gold standards based upon high-performing students offer comparable or superior predictive power relative to gold standards based upon experts.

Taken together, the results of these studies indicate that the knowledge organizations of students are qualitatively different from the knowledge organizations of experts. Experts appear to be at a qualitatively different stage or phase of learning and understanding, one which is not predictive of student performance. Expert knowledge structures may be too far removed from
student knowledge organizations to allow for great predictive power. High-performing or high-level intermediate students, however, may be at a more advanced position in the same stage of learning, allowing more powerful predictions. Thus, it appears that using a unique ideal for different stages of knowledge development improves predictive power.

Although technician #8 contributed to most or all of the predictability in the ratings and diagramming techniques, he was not predictive for two of the evaluated mental model measurement techniques. First, technician #8 was not predictive for the think-aloud technique. This lack of predictability may be more indicative of the think-aloud technique rather than using technician #8 as a gold standard. Verbalizing thoughts appeared to be problematic for the technicians. For example, some of the more experienced technicians did not verbalize, while other technicians simply read the T.O. out loud. In general, information that was verbalized was highly variable, and little of it seemed directly related to task performance. Second, technician #8 was not predictive for the laddering technique. Here, technician #8 did not perform differently from the other high performers. Perhaps the laddering technique taps into very basic mental model knowledge about existing components. This basic knowledge may not evolve with expertise to the same extent as does knowledge about component interrelations which is tapped by the ratings and diagramming techniques.

**Providing a Context.** A second interesting observation deals with the context provided during mental model measurement. Specifically, providing a troubleshooting problem context when measuring mental model knowledge produced conflicting results. For the laddering technique, providing a problem context was useful. Predicting troubleshooting performance was greatest when components relevant to the troubleshooting problem were listed (Step 2).
However, listing all system components regardless of problem context (Step 3) was not predictive of troubleshooting performance. On the other hand, providing a problem context in the diagramming technique resulted in diagrams which were less predictive of performance than were diagrams constructed without restriction to the specific troubleshooting problem. This diagramming finding could be due to the nature of the troubleshooting problem. Specifically, there were few diagram connections that were relevant to the problem. It may also be that task order is responsible for these different problem-context effects. Context was helpful in the laddering technique where the context-specific questions preceded the general ones and not in the diagramming technique where the general questions came first. Finally, the differential context effects observed for these two measures suggests that the measures may be accessing different aspects of mental model knowledge, one which is sensitive to context and one which is not. Once again, a characterization of the types of information provided by the various mental model measures would be useful. Here, it may provide some insight into the reasons for these context effects.

**Development of Expertise.** Finally, this research appears to support a pattern of development in troubleshooting expertise observed by Cooke and Rowe (1993) who examined the actions taken by technicians tasked with troubleshooting complex avionics equipment. They found that, after training, low performers exhibited a wide range of troubleshooting actions. A portion of these actions were executed by high performers, and a portion were not. These results suggest that prior to achieving expertise, but after some experience, technicians have an extensive array of knowledge in the form of executable actions, but they do not know when these actions apply. Results from the current study support this conclusion and suggest that in fact this pattern is
necessary in the development of expertise. The best technicians shared 
laddering components with the gold standard, but they also made more errors 
of commission. The fact that commission errors are predictive of performance 
indicates that the acquisition of this "extra" knowledge may, in fact, be a 
necessary stage in the development of expertise. Also, in the diagramming 
technique technicians with fewer omissions relative to the gold standard were 
more likely to be better troubleshooters. Again, the better troubleshooters may 
have a wide range of information, a subset of which is information considered 
critical by experts.

This phenomenon is also illustrated by a comparison of the network 
generated from technician #8's ratings and the aggregate network created from 
the eight lowest performers' ratings (see Figures 6a and 6b). First, the low 
performers focused exclusively on two system components: the LRU9 and the 
LRU3. Notice that all connections involve one or both of these components. 
These two components are also central in the network of technician #8, 
however, other appropriate components (e.g., the LRU2) are also central. 
Technician #8's network also includes many more links (including many weak 
one, link weights > 3.0) than the low performers' network. Thus, with 
experience, it appears that technicians not only learn more executable actions 
(Cooke & Rowe, 1993), but technicians also appear to learn more interrelations 
among system components.

Interestingly, these results correspond with Karmiloff-Smith's (1986) 
observations of children's cognitive and linguistic development. For example, 
in her studies of children's acquisition of French grammar she noted that error 
rates change with increasing knowledge. Specifically, error rates are low 
during early acquisition (around five-years old). This correct usage is followed
by a period (between approximately five- and seven-years old) when error rates increase; grammatical output becomes markedly different from that of adult grammar. Correct usage is observed once again at approximate eight-years old. Lesgold et al. (1988) noted that radiologists varying in expertise exhibit similar error patterns. That is, on some of the film stimuli more advanced radiologists were less likely to offer correct interpretations than were less advanced radiologists. Together, these results indicate that performance may not be a monotonic function of experience.

Karmiloff-Smith (1986) developed a three-phase model of the processes underlying linguistic and cognitive development based on her observations of children. During the first phase, learners are predominantly data-driven; they are working to create a match between the evaluation of their outputs (e.g., grammar) and their representations of the ideal (adult grammar). Few errors are observed. During the second phase, learners are actively working to alter and organize their internal representations. Because the learner's focus is on internal representations rather than the external environment, errors tend to increase. When the third phase is reached, learners have well-developed internal representations. Error rates similar to those occurring during the first phase are observed, but the representations underlying responses are richer and more organized. Shuell (1990) developed a similar three-phase model, describing the processes involved in meaningful learning. Incorporating such a phase approach may enhance our understanding of the development of avionics troubleshooting expertise and the development of expertise in general.

**Conclusions**

In conclusion, this research represents a first step toward the direct evaluation of the validity of several mental model measurement methods. No
other research has directly assessed the validity of mental model measures. Here, a pragmatic approach to assessing validity was taken, with the criterion being actual performance. Other research has indirectly examined validity by determining how well a measure discriminates two or more groups varying in expertise or in type of training. The problem with such approaches is that they simply demonstrate that the measure discriminates group membership. However, the mental models that differentiate groups may not correlate with performance. In fact, researchers have noted dissociations between task performance and verbalizable knowledge about that task (e.g., Broadbent, FitzGerald, & Broadbent, 1986; Reber, 1989). Simply having the ability to categorize people may not offer information directly relevant for training interventions. For example, Schvaneveldt et al. (1985) found that expert and novice pilots held different knowledge structures, but these expert/novice differences do not directly offer information regarding pedagogical interventions.

On the other hand, using performance as the final criterion indicates whether or not the representations correlate with performance. It makes intuitive sense that information of this sort would be valuable in designing a training intervention. For example, if a particular type of mental model is associated with successful performance, then training which incorporates instruction in this type of mental model should result in performance improvements. Because the representations are correlated with performance, chances are better that an intervention based on these representations would enhance performance. However, it should be noted that these sorts of relationships are correlational, making causal inferences impossible.
This study examined technicians' mental models in a cross-sectional manner. Future research could examine the validity of mental model measures in a longitudinal sense. Performance changes that occur with increasing experience could be measured along with changes in elicited knowledge. Such a study would offer information regarding those knowledge measures which are most sensitive to changes in performance with increasing experience. A similar approach has been taken in examining changes in performance with experience which are associated with the development of procedural and declarative knowledge (Neville, 1993). Neville's research points to the multi-faceted nature of performance. For example, she found that some performance metrics correlated well with certain knowledge measures, whereas other performance indices correlated with different knowledge measures. Neville's results indicate that emphasis should be placed on selecting a performance measure which adequately represents the performance variable of interest.

Finally, although this research was undertaken primarily to address applied issues, it offers a good example of how interesting basic psychological issues can be revealed in applied settings. For example, this research raised issues pertinent to the development of expertise, indicating that the development of expertise may not be a monotonic process. At the least, the results of this research raised interesting basic psychological issues, answered practical, applied questions in the domain of avionics troubleshooting, and made a step toward the validation of mental model measurement techniques.
References


Appendix A: Questionnaire 1

Component Naming Task

(The following questions pertain to the task in which you generated the troubleshooting problem system components. These components were written on index cards and laid out in front of you as you named them.)

1. This task was: difficult--1------2-------3-------4-------5------6--easy

   Please describe why you rated the task as you did : ____________________________  
   ____________________________

2. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop this task seemed:

   different--1------2-------3-------4-------5------6--similar

   Please describe why you rated the task as you did : ____________________________  
   ____________________________

3. The range of responses that I could use to express myself in this task seemed:

   restricted--1------2-------3-------4-------5------6--broad

   Please describe why you rated the task as you did : ____________________________  
   ____________________________

4. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop, this task seemed:

   artificial--1------2-------3-------4-------5------6--realistic

5. The information gained from this task is___________for measuring your System Knowledge.

   useless--1------2-------3-------4-------5------6--useful

   Please describe why you rated the task as you did : ____________________________  
   ____________________________

6. Please write your comments here and on the back of this page:
Ratings Task
(The following questions pertain to the task in which you rated the relatedness of pairs of system components, using the computer.)

1. This task was: difficult--1-------2-------3-------4-------5-------6--easy

   Please describe why you rated the task as you did:
   ____________________________________________________________
   ____________________________________________________________

2. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop this task seemed:

   different--1-------2-------3-------4-------5-------6--similar

   Please describe why you rated the task as you did:
   ____________________________________________________________
   ____________________________________________________________

3. The range of responses that I could use to express myself in this task seemed:

   restricted--1-------2-------3-------4-------5-------6--broad

   Please describe why you rated the task as you did:
   ____________________________________________________________
   ____________________________________________________________

4. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop, this task seemed:

   artificial--1-------2-------3-------4-------5-------6--realistic

   Please describe why you rated the task as you did:
   ____________________________________________________________
   ____________________________________________________________

5. The information gained from this task is___________for measuring your System Knowledge.

   useless--1-------2-------3-------4-------5-------6--useful

   Please describe why you rated the task as you did:
   ____________________________________________________________
   ____________________________________________________________

6. Please write your comments here and on the back of this page:
Diagramming Task

(The following questions pertain to the task in which you arranged and connected a set of system components, using index cards and paper arrows.)

1. This task was: difficult--1-------2-------3-------4-------5-------6--easy

   Please describe why you rated the task as you did:
   __________________________________________________________________________
   __________________________________________________________________________

2. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop this task seemed:

   different--1-------2-------3-------4-------5-------6--similar

   Please describe why you rated the task as you did:
   __________________________________________________________________________
   __________________________________________________________________________

3. The range of responses that I could use to express myself in this task seemed:

   restricted--1-------2-------3-------4-------5-------6--broad

   Please describe why you rated the task as you did:
   __________________________________________________________________________
   __________________________________________________________________________

4. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop, this task seemed:

   artificial--1-------2-------3-------4-------5-------6--realistic

   Please describe why you rated the task as you did:
   __________________________________________________________________________
   __________________________________________________________________________

5. The information gained from this task is___________for measuring your System Knowledge.

   useless--1-------2-------3-------4-------5-------6--useful

   Please describe why you rated the task as you did:
   __________________________________________________________________________
   __________________________________________________________________________

6. Please write your comments here and on the back of this page:
**Thinking Aloud Task**
(The following questions pertain to the talking aloud we asked you to do as you were troubleshooting the problem.)

1. This task was: difficult--1-------2-------3-------4-------5-------6--easy
   Please describe why you rated the task as you did: __________________________
   _________________________________________________________________

2. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop this task seemed:
   different--1-------2-------3-------4-------5-------6--similar
   Please describe why you rated the task as you did: __________________________
   _________________________________________________________________

3. The range of responses that I could use to express myself in this task seemed:
   restricted--1-------2-------3-------4-------5-------6--broad
   Please describe why you rated the task as you did: __________________________
   _________________________________________________________________

4. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop, this task seemed:
   artificial--1-------2-------3-------4-------5-------6--realistic
   Please describe why you rated the task as you did: __________________________
   _________________________________________________________________

5. The information gained from this task is _____________ for measuring your System Knowledge.
   useless--1-------2-------3-------4-------5-------6--useful
   Please describe why you rated the task as you did: __________________________
   _________________________________________________________________

6. Please write your comments here and on the back of this page:
**Troubleshooting Task**

(The following questions pertain to the task in which we asked you to troubleshoot the RWR Wiring problem verbally, with TSgt. Kruse reporting the equipment states to you.)

1. This task was: difficult--1------2-------3-------4------5------6--easy
   
   Please describe why you rated the task as you did: __________________________
   
   ____________________________________________

2. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop this task seemed:

   different--1------2-------3-------4------5------6--similar

   Please describe why you rated the task as you did: __________________________
   
   ____________________________________________

3. The range of responses that I could use to express myself in this task seemed:

   restricted--1------2-------3-------4------5------6--broad

   Please describe why you rated the task as you did: __________________________
   
   ____________________________________________

4. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop, this task seemed:

   artificial--1------2-------3-------4------5------6--realistic

   Please describe why you rated the task as you did: __________________________
   
   ____________________________________________

5. The information gained from this task is____________for measuring your System Knowledge.

   useless--1------2-------3-------4------5------6--useful

   Please describe why you rated the task as you did: __________________________
   
   ____________________________________________

6. Please write your comments here and on the back of this page:
Explaining Troubleshooting Steps Task
(The following questions pertain to the task in which we asked you to go back through the troubleshooting problem and explain why you took the actions you took and what the corresponding equipment results meant to you.)

1. This task was: difficult--1-------2-------3-------4-------5-------6--easy
Please describe why you rated the task as you did:
________________________________________________________________________
________________________________________________________________________

2. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop this task seemed:
different--1-------2-------3-------4-------5-------6--similar
Please describe why you rated the task as you did:
________________________________________________________________________
________________________________________________________________________

3. The range of responses that I could use to express myself in this task seemed:
restricted--1-------2-------3-------4-------5-------6--broad
Please describe why you rated the task as you did:
________________________________________________________________________
________________________________________________________________________

4. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop, this task seemed:
artificial--1-------2-------3-------4-------5-------6--realistic
Please describe why you rated the task as you did:
________________________________________________________________________
________________________________________________________________________

5. The information gained from this task is____________for measuring your System Knowledge.
useless--1-------2-------3-------4-------5-------6--useful
Please describe why you rated the task as you did:
________________________________________________________________________
________________________________________________________________________

6. Please write your comments here and on the back of this page:
Appendix B: Questionnaire 2
Task Listing:

**Card Task:** Generating system components-written on index cards and laid out in front of you.

**Ratings Task:** Rating the relatedness of pairs of system components, using the computer.

**Diagramming Task:** Arranging and connecting a set of system components, using index cards and paper arrows.

**Think Aloud Task:** Talking aloud continuously while troubleshooting the problem.

**Troubleshooting Task:** Troubleshooting the RWR Wiring problem verbally, specifying your actions with Sgt. Kane or Sgt. Kruse reporting the equipment states.

**Explaining Troubleshooting Steps Task:** Reviewing actions and equipment states in completed troubleshooting problem.

1. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

   A. Troubleshooting Task  
   B. Think Aloud Task

   Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

   1---------2---------3---------4---------5---------6---------7---------8
   Not at all similar                      Extremely similar

2. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

   A. Diagramming Task                     B. Think Aloud Task

   Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

   1---------2---------3---------4---------5---------6---------7---------8
   Not at all similar                      Extremely similar
3. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

A. Troubleshooting Task     B. Ratings Task

Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

1--------2--------3--------4--------5--------6--------7--------8
Not at all similar                Extremely similar

4. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

A. Diagramming Task     B. Ratings Task

Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

1--------2--------3--------4--------5--------6--------7--------8
Not at all similar                Extremely similar

5. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

A. Card Task     B. Explaining Troubleshooting Steps Task

Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

1--------2--------3--------4--------5--------6--------7--------8
Not at all similar                Extremely similar

6. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

A. Explaining Troubleshooting Steps Task     B. Ratings Task

Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

1--------2--------3--------4--------5--------6--------7--------8
Not at all similar                Extremely similar
7. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

   A. Explaining Troubleshooting Steps Task      B. Diagramming Task

   Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

   1--------2--------3--------4--------5--------6--------7--------8
     Not at all similar          Extremely similar

8. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

   A. Card Task                B. Think Aloud Task

   Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

   1--------2--------3--------4--------5--------6--------7--------8
     Not at all similar          Extremely similar

9. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

   A. Think Aloud Task          B. Explaining Troubleshooting Steps Task

   Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

   1--------2--------3--------4--------5--------6--------7--------8
     Not at all similar          Extremely similar

10. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

    A. Card Task                  B. Troubleshooting Task

    Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

    1--------2--------3--------4--------5--------6--------7--------8
     Not at all similar           Extremely similar
11. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

   A. Think Aloud Task  B. Ratings Task

   Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

   1---------2---------3---------4---------5---------6---------7---------8
   Not at all similar                          Extremely similar

12. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

   A. Card Task  B. Diagramming Task

   Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

   1---------2---------3---------4---------5---------6---------7---------8
   Not at all similar                          Extremely similar

13. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

   A. Ratings Task  B. Card Task

   Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

   1---------2---------3---------4---------5---------6---------7---------8
   Not at all similar                          Extremely similar

14. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

   A. Diagramming Task  B. Troubleshooting Task

   Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

   1---------2---------3---------4---------5---------6---------7---------8
   Not at all similar                          Extremely similar
15. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the FiWR Wiring problem in the shop.

   A. Troubleshooting Task    B. Explaining Troubleshooting Steps Task

Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

1--------2--------3--------4--------5--------6--------7--------8
   Not at all similar                        Extremely similar

16. In the space below, please write your own definition of System Knowledge:
17. We think of System Knowledge as knowledge of the components of a system, how those components fit together, and how they work together. Do you think System Knowledge is important for the actual troubleshooting of the RWR Wiring problem in the shop? Indicate your response on the rating scale below:

<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Not at all Important</td>
</tr>
<tr>
<td>2</td>
<td>Slightly Important</td>
</tr>
<tr>
<td>3</td>
<td>Important</td>
</tr>
<tr>
<td>4</td>
<td>Extremely Important</td>
</tr>
<tr>
<td>5</td>
<td>Impossible to solve problem without it</td>
</tr>
</tbody>
</table>
Appendix C: Supervisor Rating Form

Rater

Total Active Federal Military Service: _____years, _____months
Total Time in this Career Field: _____years, _____months

IDENTIFICATION OF RATEE

In this survey you are being asked to give

information about the person identified below:

________________________________________

In all the sections which follow, any reference to "this person, "this airman," the ratee," or "the person being rated" means the person identified above.
Section I: Familiarity with Ratee

For each item below, please read the question and provide the response that best describes your answer.

1. How long have you served as this person's supervisor?
   ____ years, ____ months.

2. How often do you supervise this person? (check one)
   ____ Every Day
   ____ 2-3 times/week
   ____ 1-3 times/month
   ____ less than once/month

3. How many airmen have you supervised who have the same grade and AFSC as the person being rated? (Counting this airman as "1", please check the appropriate response below.)
   ____ 1 or 2   ____ 3 to 5   ____ 6 to 10   ____ 11 or more

4. To what extent does this airman perform a "mainstream" job within his/her AFSC? In other words, to what extent does this person's job consist of tasks that are typically performed by members of this AFSC? (Check the word below which best completes the following statement)
   His/Her job contains __________ typical tasks.
   ____ no   ____ few   ____ some   ____ mostly

5. Please rate this airman according to his/her general job knowledge:
   Novice--1-----2-----3-----4-----5-----6-----7--Expert
Section II: Job Performance

In this section of the survey you will be rating the airman's job performance. You will be completing these ratings in the context of a specific troubleshooting problem. Please assume that the airman has been asked to troubleshoot the following problem within the TEWS system:

The RWR is inoperative. The BIT light is on. The TEWS display is blank.

(Connector 65P-D002H on LRU3 has one of the wires in the twisted pair shorted to the shield in the connector. This wire pair lead to the TEWS Display Unit (65P-J009A) and provides the data in the path.)

Please rate the airman's job performance in reference to this specific problem. Each statement below expresses some aspect of job performance. Use the following rating scale to indicate the extent to which you agree or disagree that the statement is an accurate description of the airman being rated. Remember, all ratings should be completed in the context of the above troubleshooting problem.

RATING SCALE FOR JOB PERFORMANCE
7 = Strongly Agree  
6 = Agree  
5 = Slightly Agree  
4 = Neither Agree nor Disagree  
3 = Slightly Disagree  
2 = Disagree  
1 = Strongly Disagree

1. Would need very little supervision in performing this assignment........

2. Would not likely need to ask others for help in performing this assignment........................................................................................................

3. Would meet local demands for speed and accuracy in carrying out this assignment...............................................................................................  

4. Would likely serve as consultant to other workers carrying out this assignment.................................................................................................................................

5. Would be capable of performing jobs other than carrying out this assignment.................................................................................................................................

6. Would carry out this assignment to the best of his/her ability...................

7. Would cooperate with supervisors & co-workers if this assignment called for teamwork ...........................................................................................................  

8. Has a strong sense of responsibility to the unit.........................................

9. Displays willingness to do more than the required amount of work......

10. Adjusts quickly and effectively to changing work situations.............
Section III: Job Knowledge

In this section of the survey you will be rating the airman's job knowledge. You will be completing these ratings in the context of the same troubleshooting problem. Again, please assume that the airman has been asked to troubleshoot the following problem:

The RWR is inoperative. The BIT light is on. The TEWS display is blank.

(Connector 65P-D002H on LRU3 has one of the wires in the twisted pair shorted to the shield in the connector. This wire pair lead to the TEWS Display Unit (65P-J009A) and provides the data in the path.)

Please rate the airman's job knowledge in reference to this specific problem. Use the following rating scale to indicate how the airman compares with other airmen in this career field who are at the same grade level. Remember, all ratings should be completed in the context of the above troubleshooting problem.

RATING SCALE FOR JOB KNOWLEDGE
7 = Very Much Above Average
6 = Above Average
5 = Slightly Above Average
4 = Average Knowledge Compared to other Airmen
3 = Slightly Below Average
2 = Below Average
1 = Very Much Below Average
X = Not Applicable to This Specialty
? = I don't know this about the airman being rated

1. Airman's knowledge of Career Development Course (CDC) material needed for this task ..............................................................

2. Airman's knowledge of essential technical procedures used in this task ...............................................................................

3. Airman's knowledge of the mission of the unit/organization...........

4. Airman's knowledge of the more difficult tasks in the specialty ....................................................................................... 

5. Airman's knowledge of the technical requirements of the specialty outside his/her present job ..............................................

6. Airman's knowledge of safety procedures pertinent to this task ..........................................................................................

7. Airman's knowledge of time-saving techniques pertinent to this task ....................................................................................

8. Airman's knowledge of required forms pertinent to this task.......
RATING SCALE FOR JOB KNOWLEDGE
7 = Very Much Above Average
6 = Above Average
5 = Slightly Above Average
4 = Average Knowledge Compared to other Airmen
3 = Slightly Below Average
2 = Below Average
1 = Very Much Below Average
X = Not Applicable to This Specialty
? = I don’t know this about the airman being rated

9. Airman’s knowledge of technical reference materials such as operating manuals, Technical Orders (TOs), or standard reference books pertinent to this task.................................................................

10. Airman’s knowledge of how to locate other people with specialized expertise pertinent to this task ................................

11. Airman’s knowledge of the proper technical terminology needed to discuss objects, methods, and goals pertinent to this task .......

12. Airman’s knowledge of the underlying principles, ideas, or concepts pertinent to this task .................................................................

13. Airman’s knowledge of what can go wrong in completing this task and how to avoid these problems ..............................................

14. Airman’s knowledge of how to cope with unexpected problems in technical assignments.................................................................

15. Airman’s knowledge of the tools and equipment pertinent to this task.......................................................................................

Appendix D: Verbal Protocol and Action Coding Scheme

**RWR Wiring Taxonomy**

+: Verbalization units used for coding the RWR Wiring--Broken Wire problem
#
: Action units used for coding the RWR Wiring--Broken Wire problem

**Debrief Questions**
1.0 Ask pilot: Have an ECS light? #
2.0 Ask pilot: Any indications on MCCP when turned RWR on? #
3.0 Ask pilot: Any indications on TEWS scope when turned RWR on? #
4.0 Ask pilot: Did malfunction exist during entire flight? #
5.0 Ask pilot: Did TEWS scope flicker on at any time? #
6.0 Ask pilot: Everything O.K. before flight? #
7.0 Ask pilot: Cycle RWR off/on? #
8.0 Ask pilot: Check ASP panel? #
9.0 Ask pilot: Do you have an AI/SAM light now? #
10.0 Ask pilot: Did you ever have an AI/SAM light? #
11.0 Ask pilot: Was the AI/SAM light on hard? #
12.0 Ask pilot: Did you have a flashing AI/SAM light? #
13.0 Ask pilot: Has RWR been on for awhile and if so is it still on? #
14.0 Ask pilot: Did you do a BIT check? #
15.0 Ask pilot: Was the BIT light on steady the whole time? #
16.0 Ask pilot: Did you have a BIT light? #
17.0 Ask pilot: Did you try to defeat the tones? #
18.0 What is my fault reporting code? #

**Visual Inspections**
19.0 Check ASP panel #
20.0 Check LRU Latches
   20.1 Check problem-appropriate LRU latches #
      20.1.1 LRU2 latches (under Door 6-R)
         20.1.1.1 Prior to letting RWR run
         20.1.1.2 After letting RWR run for approx. 15-30 min.
      20.1.2 LRU3 latches (under Door 6-R)
         20.1.2.1 Prior to letting RWR run
         20.1.2.2 After letting RWR run for approx. 15-30 min.
20.1.3 LRU9 latches  
  20.1.3.1 Prior to letting RWR run  
  20.1.3.2 After letting RWR run for approx. 15-30 min.  

20.2 Check problem-inappropriate LRU latches #  

20.2.1 LRU6 latches  
  20.2.1.1 Prior to letting RWR run  
  20.2.1.2 After letting RWR run for approx. 15-30 min.  

21.0 Check RWR circuit breakers #  

22.0 Check LRU Connections  

22.1 Check problem-inappropriate LRU Connections #  
  22.1.1 LRU2 connections  
    22.1.1.1 Cannon plugs  
    22.1.1.2 Pins  
    22.1.1.3 ARF connectors  
  22.1.2 LRU3 connections  
    22.1.2.1 Cannon plugs  
    22.1.2.2 Pins  
    22.1.2.3 ARF connectors  
  22.1.3 LRU9 connections  
    22.1.3.1 Cannon plugs  
    22.1.3.2 Pins  
    22.1.3.3 ARF connectors  

22.2 Check problem-inappropriate LRU Connections #  
  22.2.1 LRU6 connections  
    22.2.1.1 Cannon plugs  
    22.2.1.2 Pins  
    22.2.1.3 ARF connectors  

23.0 Visually check aircraft wiring #  
  23.1 Pull cannon plug off & visually check wires 65P-D002H-65P-D009A.  
  23.2 Visually check for chaff  
    23.2.1 65P-D002H-65P-D009A.  
    23.2.1.1 from LRU3 end.  
    23.2.1.2 from LRU9 end.  
  23.3 Visually check pins
24.0 Check LRU2 fuses #
25.0 Check LRU ETI meters
   25.1 Easy Check: ETI meter #
      25.1.1 LRU2 ETI meter
      25.1.2 LRU3 ETI meter
   25.2 Difficult Check: ETI meter #
      25.2.1 LRU6 ETI meter
      25.2.2 LRU9 ETI meter
26.0 Turn control panel off/on; Cycle RWR off/on. #
27.0 Turn RWR on & check if BIT light flashes. #
28.0 Is the RWR on/off switch in "on" position? #
29.0 Turn RWR on--does AI/SAM light flash? #
30.0 Is A/I SAM light on or off, steady? #
31.0 Adjust TEWS scope intensity to full bright. #
32.0 Did EWW light go off when turned on? #
33.0 Do fault indicators reset? #
73.0 Turn RWR off--Do latches remain latched?#

**Built-in Tests (BIT)**
34.0 RWR BIT check (duplicate symptoms?) #
35.0 RWR BIT check--let RWR run for > 5 minutes (duplicate symptoms?) #
36.0 Other system BIT #
   36.1 IBS (Interference Blanker) BIT
   36.2 CC (Central computer) BIT

**Reprogram LRU**
37.0 Reprogram LRU3. #

**Audio Inspections**
38.0 Check: Getting audio tones? #

**Aircraft History**
39.0 Check forms: Is this a repeat failure? #
Swaps
40.0 Swap LRU:
   40.1 Swap Problem-appropriate LRU #
      40.1.1 Swap LRU2
      40.1.2 Swap LRU3
      40.1.3 Swap LRU9
   40.2 Swap Problem-inappropriate LRU #
      40.2.1 Swap LRU6
      40.2.2 Swap LRU10
      40.2.3 Swap LRU11
41.0 Swap ASP panel #
42.0 Swap (or repair) video cable #
43.0 Swap wire
   43.1 65P-J009C #
   43.2 65P-J009A #
44.0 Swap BIT Control Panel #
45.0 Swap connector
   45.1 On LRU9 side. #
   45.2 On LRU3 side. #

Measurements: Wire Continuity
46.0 Measure wire continuity LRU2--LRU3 #
   46.1 65P-D001G (LRU2) to 65P-D002L (LRU3)
      46.1.1 all pins to all pins
      46.2.2 pin 67 and pin 67.
      46.2.3 pin 127 and pin 127.
   46.2 from 65P-D001E (LRU2) to 65P-D002D (LRU3)
      46.2.1 pin 27 to pin 27
      46.2.2 pin 28 to pin 28
      46.2.3 pin 30 to pin 30
      46.2.4 pin 31 to pin 31
      46.2.5 pin 127 (LRU2) to pin 126 (LRU3).
47.0 Measure wire continuity LRU6--LRU3 #
   47.1 65P-T005G (LRU6) to 65P-D002A (LRU3)
   47.2 5J7 (LRU6) to 2J1 (LRU3)
47.3 5J11 (LRU6) to 2J5 (LRU3)
47.4 5J6 (LRU6) to 2J7 (LRU3)
48.0 Measure wire continuity LRU6--LRU2 #
48.1 5J1 (LRU6) to 1J3 (LRU2)
49.0 Measure wire continuity LRU2--LRU9 #
49.1 All wires, all voltages from LRU2 to LRU9.
49.2 65P-J001B to 65P-D009B
   49.2.1 pin 53 to pin 53
   49.2.2 pin 25 to pin 25 for 28 Volts
   49.2.3 pin 54 to pin 54 for 115 Volts AC
   49.2.4 pin 19 to pin 19 for -25 Volts.
   49.2.5 pin 17 to pin 17 for 100 Volts.
   49.2.6 pin 23 to pin 23 for -5 Volts.
   49.2.7 pin 27 to pin 27.
   49.2.8 pin 28 to pin 28.
   49.2.9 pin 29 to pin 29.
   49.2.10 pin 30 to pin 30.
   49.2.11 pin 31 to pin 31.
49.3 1J2 (LRU2) to 9J2 (LRU9)
   49.3.1 end-to-end
   49.3.2 shorts between wires
50.0 Measure wire continuity Circuit Breaker Panel--LRU2 #
50.1 Circuit Breaker Panel to LRU2, 1J1
   50.1.1 end-to-end
   50.1.2 each wire to A/C ground
51.0 Measure wire continuity--Lights Test Relay Panel #
51.1 Through Lights Test Relay Panel
51.2 pins 4 to 27
52.0 Measure wire continuity--Miscellaneous #
52.1 Measure wire continuity from 65P-D002B (LRU3), pins 11 and 13 to
   pin 25 and A/C grnd. 52.2 Place BCP selection switch to RWR and
   press initiate.
52.3 TEWS circuitry
52.4 Power wires coming from LRU3
53.0 Measure wire continuity LRU3--LRU9
53.1 Between 65P-D002H and 65P-J009A

53.1.1 Center conductor to shield #
   53.1.1.1 Center conductor to shield, someone at other end shorting shield.
   53.1.1.2 Pin 1 to shield #
   53.1.1.3 Pin 2 to shield #

53.1.2 Pin 1 to Pin 2 #
   53.1.2.1 on LRU3 side
   53.1.2.2 on LRU9 side
   53.1.2.3 Jump from center conductor (pin 1) to outside conductor (pin 2).

53.1.3 Center pin to A/C ground #
   53.1.3.1 Pin 1 to A/C ground
      53.1.3.1.1 With ground (shorting wire 1 to ground)
      53.1.3.1.2 With no ground.
      53.1.3.1.3 Pin 1 to ground at each end.
   53.1.3.2 Pin 2 to A/C ground.
      53.1.3.2.1 With ground (shorting wire 1 to ground)
      53.1.3.2.2 With no ground.
      53.1.3.2.3 Pin 2 to ground at each end.

53.1.4 Center pins: End-to-end #
   53.1.4.1 Pin 1 to Pin 1 (end-to-end)
   53.1.4.2 Pin 2 to Pin 2 (end-to-end)

53.1.5 Shield to A/C ground #
   53.1.5.1 With ground (shorting wire 1 to ground)
   53.1.5.2 Shield to A/C ground, no ground.
   53.1.5.3 Pin 1 to A/C gnd.
      53.1.5.3.1 on LRU3 side
      53.1.5.3.2 on LRU9 side

53.1.6 Shield to shield #

53.1.7 Remove LRU3 connector and shoot bare pins, bare wires#

53.2 Between 65P-D002J and 65P-J009C (off the active path) #

53.2.1 Pin to pin
   53.2.1.1 Pin 3 to Pin 4
   53.2.1.2 On 65P-J009C: jumper both pins from LRU3 back to
53.2.2 Center conductor to shield.
   53.2.2.1 Pin 3 to shield
   53.2.2.2 Pin 4 to shield
   53.2.3 Center pins: End-to-end
   53.2.3.1 Pin 3 to Pin 3
   53.2.3.2 Pin 4 to Pin 4
   53.2.4 Shield to shield
   53.2.5 Center conductor to ground
      53.2.5.1 Pin 3 to ground at each end
      53.2.5.2 Pin 4 to ground at each end
53.3 Between 65P-D001F and 65P-J002P (off the active path) #
   53.3.1 Pin 4 to pin 4.
   53.3.2 Pin 5 to pin 5.
   53.3.3 Pin 1 to pin 1.
   53.3.4 Pin 2 to pin 2.
54.0 Measure wire continuity TEWS Control Panel--LRU3 #
   54.1 pin 1 to pin 1
55.0 Measure wire continuity TEWS Control Panel--Miscellaneous #
   55.1 from 65P-J010 (TEWS CP), pin 2 to Avionics Protection Relay No.5, pin 4
   55.2 from 65P-J010 (TEWS CP), pin 2 to 52P-J083B (Light Test Relay Panel), pin 4.
56.0 Measure wire continuity Light Test Relay Panel--LRU3 #
   56.1 pin 27 to pin 2.
57.0 Measure wire continuity Data Processor--Input Data Link #

Measurements: Voltage/Power
58.0 Measure Voltage/Power to the LRU2 #
   58.1 28 Volts DC on 1J2, pin 25 to 32.
   58.2 Check pin 5--getting 28 Volts DC to LRU2?
59.0 Measure Voltage/Power--General #
   59.1 Is the system getting power?
Verbalizations
60.0 Action interpretations
   60.1 Power +
   60.2 Duplicate fault +
   60.3 Connections +
   60.4 Ease of Action +
   60.5 T.O. Says... +
   60.6 Previous result +
   60.7 Latch +
   60.8 Information Flow +
   60.9 Display +
   60.10 Tones +
   60.11 Shorts +
   60.12 Continuity +
   60.13 AI/SAM Light +
   60.14 BIT light +
   60.15 Solve problem/solution +
   60.16 Bad parts +
   60.17 Miscellaneous/Frustration +
61.0 Result interpretation +
62.0 Elimination +
63.0 Elimination Justification +
64.0 Plan/prepare for test/check +
65.0 T.O. Search/interpretation +