DOCUMENT RESUME

ED 300 397 TM 012 372

AUTHOR	Hambleton, Ronald K.; And Others
TITLE	New Testing Methods to Assess Technical Problem-Solving Ability.
SPONS AGENCY	Air Force Human Resources Lab., Alexandria, Va. Manpower Development Div.; Human Resources Research Organization, Alexandria, Va.
PUB DATE	15 May 88
NOTE	45p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988). Figure 2 contains small/marginally legible print.
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Achievement Tests; Aircraft Pilots; *Computer Assisted Testing; *Occupational Tests; Personnel Evaluation; *Problem Solving; *Technical Occupations; *Test Construction; Testing; Test Validity
IDENTIFIERS	*Air Force; Diagnostic Achievement Tests

ABSTRACT

Tests to assess problem-solving ability being provided for the Air Force are described, and some details on the development and validation of these computer-administered diagnostic achievement tests are discussed. Three measurement approaches were employed: (1) sequential problem solving; (2) context-free assessment of fundamental skills and knowledge; and (3) constrained tasks. Computer administration was chosen for its advantages of: (1) dynamic control; (2) variable response mode; (3) capturing responses and scoring; and (4) test security. Disadvantages were considered to be outweighed by these aspects. The Diagnostic Achievement Tests consist of Part I, a computerized sequential problems test, and Part II, the Enabling Skills Test (an independent assessment of each skill). Subtests are discussed briefly, and the 15-step process of test development is summarized. Special problems in the development of these specific tests are discussed, including excessive test length, glitches in the computer software, and flaws in the test orientation and approach to obtaining airmen's hypotheses about the locations of faults as they work through the problems. Since initial pilot testing was informative and supportive, validity studies will be done. Four figures support the process description. (SLD)

New Testing Methods to Assess Technical Problem-Solving Ability

Ronald K. Hambleton, Charlene Gower, John Bollwark
University of Massachusetts at Amherst

5/15/88

New Testing Methods to Assess Technical Problem-Solving Ability[1,2]

Ronald K. Hambleton, Charlene Gower, John Bollwark
University of Massachusetts at Amherst

The research described in this paper is part of a large Air Force project designed to improve the job effectiveness of airmen through improved training of problem-solving skills (Gott, 1988; Hambleton, 1986). Specifically, interest in the project is centered on problem-solving skills that (1) are measurable, (2) are trainable, and (3) are useful in distinguishing expert from novice performers.

The nature of the skills to be measured in the project was such that standard achievement testing methods were judged as unsuitable for producing valid measurements. For example, the multiple-choice format is too limited to handle the situation where several of the answer choices to a question may be correct or must be rank-ordered by examinees. Cuing of correct answers is also a shortcoming of the multiple-choice format.

In developing valid diagnostic tests of problem-solving skills needed for successful performance in the electronics Air Force specialties, our view was that the tests would need to have certain characteristics: First, it seemed essential to build the tests around technical problems that arise in the Air Force specialties of interest.

---

[1]Paper presented at the annual meeting of AERA, New Orleans, April, 1988.

[2]The University of Massachusetts is one of several subcontractors to the Human Resources Research Organization (HumRRO) on a five-year contract with the Air Force (Dr. Sherrie P. Gott, Project Monitor) entitled, "Development of an Integrated System to Assess and Enhance Basic Job Skills."

Labrepor.1.1

In this way, the problem-solving skills could be assessed in an appropriate job context, rather than in isolation. The generalizability of the test score interpretations would certainly be enhanced if the skills were assessed in a job-related context. Second, it seemed that several new item formats would be needed to increase the validity of the test scores. The multiple-choice format and related objective formats were viewed as too limiting to facilitate the assessment of many of the cognitive skills of interest. Third, obtaining valid measurements seemed to require the development of tests that would allow airmen to solve problems in much the same way they would attempt to solve them on the job. Clearly, then, tests would need to be highly adaptive to the problem-solving preferences of airmen. Also, the order of presentation of test material would need to be unique for each airman and be dependent upon his/her preferences and performance during the test. Finally, such flexibility in test question sequencing seemed to require the aid of microcomputers for test administration. A manually administered adaptive test would be cumbersome and reduce flexibility in comparison to the flexibility offered by microcomputers (see, for example, Nitko & Hsu, 1984).

The main purpose of this paper is to provide an overall description of the tests being developed for the Air Force and some of the details concerning the development and validation of these new computer-administered diagnostic achievement tests to measure problem-solving skills. The tests have several interesting features which are highlighted in the next sections.

Background Issues

Cognitive Variables to Measure

The taxonomy of skills of interest in the project was divided into four branches: (a) declarative knowledge, (b) procedural skills, (c) procedural problem-solving operations, and (d) metacognitive skills. Figure 1 provides a description of the relationship among the four branches. Declarative knowledge is an understanding of how, for example, an electronic computer or radar system works. The declarative knowledge branch involves component knowledge and system knowledge.

The procedural skills branch involves knowledge and skills in the methods employed in accomplishing the task of problem solving. Within an Air Force Specialty (AFS) related to the maintenance and calibration of electronic, computer, and radar systems, for example, procedural skills involve the steps to follow to identify, test, repair, and calibrate electronic, computer, and radar systems and subsystems. Such knowledge and skills can be described as basic operations and intermediate operations.

Both the procedural skills branch and the declarative knowledge branch are fully realized in procedural problem-solving operations. Here, we are looking for problem-solving skills such as planning or space splitting as they are applied to troubleshooting for a particular equipment system. All of these operations are embedded in the problems that are identified in the task analysis phase of the project and will represent the important cognitive skills that are of interest. The problems focus on common, albeit difficult, tasks and include multiple significant occurrences of the cognitive skills.

Metacognitive skills can be loosely defined as being aware of one's thinking processes and knowledge (Sternberg, 1985). Experts differ from novices in two important respects: (1) experts are more exhaustive in their use of available information to solve a problem, and (2) experts spend more time planning how to go about solving a problem and less time actually doing the "solving."

Measurement Strategies

Based upon our work with the categories of skills shown in Figure 1, three different measurement approaches seemed necessary: (a) sequential problem solving, (b) context-free assessment of fundamental skills and knowledge, and (c) constrained tasks.

The sequential problem solving takes the form of complex sequential branching problems, where the branches taken by the airmen depend upon their prior responses. These tests simulate the actual decisions and activities of, for example, troubleshooting a faulty test station switching complex. Critical procedural skills (such as running a serial loop test), problem-solving strategy (such as using a method to check one's work), and other critical skills can be assessed within this problem simulation. In this way, the particular cognitive skills of interest to the project could be measured within the context of realistic problems.

The context-free assessment of fundamental skills and knowledge is primarily focused on the procedural skills and declarative knowledge categories. These tests measure an airman's understanding and mastery of fundamental skills. The skills are not measured in a complex problem context. One advantage of such general-content context-free items is that they could be used for assessment in more than one Air Force specialty.

The constrained task approach is focused on the assessment of skills at the intermediate operations and systems knowledge level. This approach takes the form of a brief presentation of a problem context followed by questions such as "what is the next step?", "how do these parts relate to each other?", and "what kind of meter reading should be expected?"

The latter two testing strategies (context-free assessment and constrained tasks) offer valuable information about airmen who fail to reach appropriate solutions to the sequential problems test. That is, if an airman fails the sequential problems test, the failure can be attributed to (1) a deficiency in the problem-solving skills required for problem success, (2) inadequacy in the supporting knowledge and skills base, (3) inability to "orchestrate" the simultaneous application of the multiple skills needed for problem solution, or (4) inability to make use of existing knowledge in appropriate situations. With the measurements obtained from the latter two types of measures, the ambiguity concerning reasons for failure on the sequential problem tests can be reduced.

Computer-Administered Tests

Microcomputers are receiving wide use in instruction and, to a lesser extent, use in item banking and test development (Nitko & Hsu, 1984). However, their use to date in administering tests and scoring examinees has been limited. There appeared to be four advantages to computer-administered tests in our research:

<u>Dynamic Control</u>. In traditional paper-and-pencil testing, the control of what questions an airman will encounter is dependent upon (a) the built-in order of questions and possible branching from answer choices to the next question, and (b) the behavior of the airman reflected in the answers chosen, the questions skipped, and the speed and continuance of working. Computer-administered tests can allow for control over the choices of questions to administer. Computers can be programmed to consider multiple patterns of responses to determine the future order of presentation of questions. In addition, decision rules can govern the immediate and future status of the testing.

<u>Variable Response Mode</u>. Computers can be used to ask a wider variety of questions and responses than traditional paper-and-pencil tests. For example, computer-administered tests can be in free response mode (where the airman furnishes the answer) when the possible answers are of a known limit, as in providing a numeric answer to a question or in identifying the appropriate, specific name of a component part. In addition, computers can easily handle questions that (a) require multiple responses or (b) allow for more than one correct answer.

<u>Capturing Responses and Scoring</u>. Because the locus of control in a traditional paper-and-pencil test is vested with the airman's dynamic performance and the static structure of the test, certain important testing variables cannot be easily measured (such as response latency)

and scoring generally occurs after the testing session. Computers, because of their ability to monitor and process data dynamically during the testing, can capture such variables as response latency and can offer rapid scoring for examinee feedback. In addition, a test can be administered any time the terminal is available and without the presence of an examiner.

Test Security. Because of the dynamic control possible with computer-administered tests, greater test security can be obtained. Examinees' reference to previous or prior test questions can be controlled so such information does not interfere with the performance. This is a particularly important advantage in presenting sequential problem tests where airmen may be tempted to look ahead in order to find a "backwards" solution to the problems. Literally thousands of sequences for taking the test can be easily accommodated with computer-adaptive testing.

But there are several disadvantages, too:

Novel Administration. Because of the "newness" of computer-administered tests, the novelty of the administration may interfere with the airman's performance on the test. Clearly, familiarity with the situation and an understanding of the testing procedures would help alleviate this disadvantage. Materials that introduce the testing strategy to an examinee, train the examinee in the necessary control of the computer, and offer sufficient practice, must be included in implementation of computer-adaptive testing.

Equipment Compatibility. There is a wide variety of computer equipment, operating systems, and software. Thus, the generalizability of a particular computer testing program is likely to be limited at least within compatible machines.

Equipment Reliability. Although computers tend to be highly reliable, at some time the computer used for testing is going to fail. When the equipment fails, the testing stops. This problem requires the maintenance of a supplemental computer that can be used to back up the testing stations and the provision for rapid repair of broken equipment.

Software Availability. Many software packages are available for developing and administering tests. Most of these packages, however, restrict the test developer to using standard item formats such as matching and multiple choice. More complex sequential problem tests place even greater demands on a software package.

Software Reliability. Software, like computers, tends to be highly reliable. However, software failures will at some time occur. Software that has been "modified" to meet certain testing needs will also be more susceptible to failure.

Software Setup Time. Tests that make use of linear branching do not require much software setup time. Non-linear branching tests, on the other hand, require individual frame branching definitions which is a time-consuming process.

Software and Computer Costs. Compared to the development of traditional paper-and-pencil tests, computer-administered tests are more expensive to develop. The additional cost of software and computer hardware may be substantial.

Screen Limitations. Examinees may only view one monitor screen's worth of information at any one time. This situation leaves the test developer with three options: (a) fit the complete item on a single screen, (b) use a scrolling item feature, or (c) make "extra" written material available. For longer items, each of these options has drawbacks. For example, fitting the item on a single screen may make the item difficult to read or confusing.

General Computer Use Problems. There are a number of general problems that arise in computer-administered tests: power failures, backing up/restoring mistakes, problems with transporting and using floppy disks, and no software system is absolutely foolproof to user errors which can shut down the system.

Our view was that the disadvantages with computer-administered tests could be overcome, albeit with difficulty, and the advantages were so important to the success of our tests that computers would need to become an integral part of the test administration. Currently, the diagnostic achievement tests are administered on a Zenith-248 microcomputer.

Test Description

The Diagnostic Achievement Tests (DAT) consist of two parts. Part I of the typical diagnostic achievement test, the Computerized Sequential Problems Test (CSPT), consists of (approximately) four problems to solve with each problem requiring about 20 to 30 steps. Cognitive skills are assessed in the presence of other skills through job-relevant problems. Since many points within each problem require offering the examinee multiple responses, the branching capability is essential, and item scoring is complex. In addition to cognitive skill scores, examinees are all assigned "overall scores" to reflect their problem-solving efficiency (percent of positive acts they took during the test), and proficiency (a percent total score reflecting the level of correctness of their answers of the maximum possible score). Some of the cognitive skill scores produced from airmen responses to the DAT questions will be useful to airmen in diagnosing their own strengths and weaknesses; other scores will be useful to trainers.

One of the additional unique features of the CSPT is that airmen must continuously update a working hypothesis list which is scored to reflect how airmen use the information they are given to solve problems.

In Part II of the tests, each cognitive skill is assessed in isolation from other skills, though job-relevant stimuli are used in the test item stems to enhance test relevance, job relatedness, and validity. Part II, the Enabling Skills Test (EST), contains (mainly) objectively scored questions to measure basic and intermediate operations and component and system knowledge. These are both the general context-free and constrained task measurement approaches described earlier. The Part II test has other characteristics: branching is not necessary, and objectively scored test items predominate. Although Part II of the tests can be administered in a test booklet format (with a separate machine-scorable answer sheet), computer terminals are used for reasons of consistency of format (since Part I requires computer delivery) and convenience in scoring.

The computer is used to administer and record scores on Part I. The essential TOs, schematics, job aids, and other job-relevant material required for the test problems are provided in a booklet that can be easily accessed and used during the Part I test. A summary of the main characteristics of the DAT (CSPT and EST) is contained below:

<u>CSPT and EST</u>

- Administered at a computer terminal.
- Total testing time is between 300 and 360 minutes.
- Cognitive skills measured in the DAT are identified in a cognitive task analysis (Task 1 in the project).
- An effort is made to measure each skill with several test items (four test items if the scoring is dichotomous).
- As many relevant skills as possible are measured within a job-relevant context (i.e., in the CSPT).

<u>CSPT</u>

- Consists of job-relevant problems.
- Job-relevant material (schematics, computer code, etc.) is presented in a separate test booklet.

- o Branching is used in the test to follow up particular airman responses.

EST

- o Includes questions about essential or enabling job knowledge and skills.
- o Skills measured may be basic to many AFS.
- o Skills are assessed in a somewhat job-independent way (to facilitate the uses of the ESTs across several AFS).
- o A linear sequence of test questions is used.

Each subtest in the DAT will be briefly discussed next.

I Computerized Sequential Problems Test (CSPT)

This subtest is the longest, and involves solving four technical problems like those that airmen work on in the specialty. Through the context of these simulated problems, a variety of procedures, strategic and metacognitive skills can be assessed. Based upon approaches to solving the problems, skills can be assessed, and scores can be produced that reveal strengths and weaknesses.

Each problem in the CSPT has three main parts:

A. Problem Statement
B. Hypothesis List
C. Action Steps

Each part will be described briefly next.

A. Problem Statement

Each problem in the CSPT begins with a problem description. An example is offered below:

* *

Problem Statement

While running an IRE LRU on the display test station, the test station indicates a fail at test number 25428 with the following printout: UUT failed test 2545; TO 12P4-2APX-218-1.

* *

The airmen are also provided a paper copy of the Problem Statements so that they may refresh their memory about the problems when needed.

B. Hypothesis List

In order to monitor an airman's thinking as he/she progresses through a problem and to assess specific problem-solving skills (e.g., ability to constrain hypotheses), the test taker is often asked to indicate all of the locations which he/she thinks could contain the fault. Since the suspected areas will change as the airman gathers information, a new hypothesis list is completed after receiving new information.

The hypothesis list is presented in a series of frames with the directions:

* *

Mark the areas you suspect with an "X" or a "P".

"X" indicates that you suspect the area, but you have no idea what locations within the area could be at fault.

"P" indicates that you suspect one or more locations within the area. Areas that are not suspected should be left blank.

* *

This differential manner of marking suspected areas allows airmen to relate their more specific hypotheses without the annoyance of being asked for detailed information they are not yet prepared to give.

Suppose that an airman thinks the fault could be one of the cards in the ECP or something in the Pulse Generator. Frames A, B, and C below show the series of frames the airman would see. Notice the manner in which the response "P" determines what subsequent frames are presented. Since Frame C is the final level of the Hypothesis List, "P" is no longer an answer choice.

* *

Frame A

Mark the areas you suspect with an "X" or "P".

"X" indicates that you suspect the area, but you have no idea what locations within the area could be at fault.

"P" indicates that you suspect one or more locations within the area. Areas that are not suspected should be left blank.

a. ____ LRU
b. ____ Test Package
c. _P__ Test Station

* *

* *

Frame B

Mark the areas you suspect with an "X" or "P".

"X" indicates that you suspect the drawer, [in the test station] but you have no idea what locations within the drawer could be at fault.

"P" indicates that you suspect one or more areas within the drawer. Drawers that are not suspected should be left blank.

a.	P	Pulse Generator	b.	____	CCDP
c.	____	Printer I/O	d.	____	AUX B
e.	____	DMM	f.	____	Frequency Counter
g.	____	SWDS	h.	X	ECP
i.	____	Video Unit	j.	____	Display Monitor
k.	____	LRUPS	l.	____	ACRPS

* *

* *

Frame C

Type an "X" beside the Pulse Generator Components you suspect.

a.	____	A13 Card	b.	____	A10 Card
c.	____	A5 Card	d.	X	A11 Card
e.	____	A7 Card	f.	____	A6 Card
g.	X	A1 Card			

* *

In summary, the Hypothesis List consists of up to three levels:

1. For problems where an LRU, Test Package, and Test Station are involved, the first level of the hypothesis list will ask for broad statements about the location of faults.

2. The second level of the hypothesis list will ask for more detailed hypotheses for any options in the first level that were designated with a "P." That is, airmen are asked to identify suspected parts of the LRU or Test Package, or suspected drawers in the Test Station.

3. The third level of the hypothesis list allows airmen to identify suspected areas of Test Station drawers for those drawers designated with a "P" in the second level.

C. Action Steps

In order to troubleshoot the problems in the CSPT, airmen have the opportunity to gather information much as they do on the job. Airmen see a list of action steps from which they choose a step to gather pertinent information. Sometimes a series of questions must be answered to indicate the specific action an airman wants to take.

Frames D through G present an example of the action steps part of the CSPT. If an airman wished to swap the ECP drawer in the test station, the airman would have chosen response "d" on Frame D, response "c" on Frame E, and response "j" on Frame F. The next screen viewed (Frame G) would give the results of the airman's chosen action, swapping the ECP.

* *

Frame D

1. What is the next step you would take to locate the problem?

   a. Take a measurement
   b. Run a programmed test
   c. Use the ECP
   d. Swap or replace a piece of equipment
   e. Check front panel controls and indicators
   f. Check fuses or major components
   g. Inspect cable connection(s)
   h. Recycle station power

(choice: d)

* *

Labrepor.1.16

* *

Frame E

2. What type of equipment do you want to swap or replace?

   a. Interface Adaptor
   b. LRU
   c. a drawer
   d. a component of a drawer
   e. a test package cable
   f. an internal test station cable
   g. a component of a cable
   h. sampling head
   i. an overhead cable

(choice: c)

* *

* *

Frame F

3. What drawer do you want to swap or replace?

   a. Switching Complex
   b. AUX A
   c. CCDP
   d. Printer I/O
   e. AUX B
   f. DMM
   g. PDP
   h. Frequency Counter
   i. O-Scope
   j. ECP
   k. Pulse Generator
   l. Data Coupler
   m. Sampling Analyzer
   n. LRUPS

(choice: j)

* *

* *

Frame G

RESULT: Swapping or replacing the ECP does not solve the problem. Problem symptoms remain the same.

* *

The three parts of the CSPT (Problem Statement, Hypothesis List, Action Steps) interact in a logical fashion. First, the Problem Statement introduces the initial problem symptoms. Then the testing process alternates between the Hypothesis List and the Action Steps, allowing the airman to repeatedly report suspicions and gather new information, until the problem is completed. After receiving a message that the problem is completed, the airman begins a new problem by viewing a new Problem Statement.

II. Enabling Skills Test (EST)

The EST, presented after the CSPT, is primarily focused on the assessment of procedural skills and declarative knowledge. It can be loosely viewed as tapping the more basic skills which may help to explain failures in the more integrated CSPT.

In the EST, approximately 30 job-relevant questions (with about 80 scorable units) address specific skills that are prerequisite to being a good trouble-shooter. A wide variety of item formats was utilized in assessing this range of skills. All items in the EST are presented at a computer terminal and are objectively scored.

Specific Test Development Steps

The development of the DATs is based on protocol analyses of airmen verbally troubleshooting technical job-related problems. These protocol analyses are complete descriptions and skill breakdowns of an airman's troubleshooting steps.

Included in each protocol analysis[1] are:

(1) Problem Overview - An introductory description of the technical problem the airman is asked to solve.

(2) Complete Problem Representation - A diagram of equipment, signal values and directions of signal flow related to this problem.

(3) Problem Generation Protocol - Complete transcripts of the problem generation session.

---

[1]Products from the protocal analysis have changed over the course of the project to meet specific needs of users. The current list of products was used in our first test development effort.

(4) Problem Generation List Protocol - A 3- to 4-page summary of the problem generation session.

(5) Solution Protocols - Completed individual transcripts of novice, mid-level, and expert airmen troubleshooting the same problem.

(6) Solution List Protocol - A 4- to 5-page summary of each of the novice, mid-level, and expert solution protocols.

(7) Solution Path Graph - A graph of the troubleshooting steps and subsequent conclusions for each of the novice, mid-level, and expert solution protocols.

(8) Effective Problem Space Graph - A composite graph of the solution Path Graph described above.

(9) Skills Analysis Graph - An Effective Problem Space Graph with corresponding skills labeled for a specific action or set of actions.

(10) Summary of Expert-Novice Skill Difference - A comparison of the different troubleshooting steps, skills used, and underlying problem representation of the novice, mid-level, and expert airman while solving this problem.

A total of (about) twelve protocol analyses are developed for each Air Force specialty. Of the twelve available problems, a set of four representative problems are chosen for the development of the computer-administered sequential problem solving test (CSPT). The remaining protocol analyses are used as a basis for developing fundamental skill and constrained task items.

Additional materials used in test development include a set of skills to be measured, corresponding skill definitions, a skills by problem matrix, and reports on technical and problem-solving issues. A final and invaluable source of help with technical details was a group of Air Force subject matter experts.

The initial protocol analysis products we worked with are based on the F-15 Electronics Maintenance specialty. Airmen in this specialty are responsible for the repair of sophisticated jet airplane electronic systems. Test equipment includes wall-sized testing stations containing thousands of components. Related technical orders (schematics and computer code) are, of necessity, quite voluminous.

The following list provides a summary of the 15-step process used in the development of the present form of the DAT.

<u>Step 1</u> - Sort through the 12 available protocol analyses to identify a subset of four problems to be used in the sequential problems test. The selection criteria included choosing problems that:

a) Cover important and hard, but not uncommon, types of problems for novice airmen. Of interest are problems that novices as well as experts attempt to solve on the job.

b) Represent a wide array of important and higher-order cognitive skills. Figure 2 provides a list of skills needed to solve each problem. Material displayed in Figure 2 is used in selecting test problems.

c) Are representative of the three main categories of problems in the specialty. (In the case of the AFS 326X4B, these are signal flow, data flow, and power.)

d) Appear to be especially interesting to airmen.

e) Lend themselves to assessment via computer-administered tests (for example, problems that require more than one airman to work on at a time would not be of interest).

f) Are technically unambiguous from the point of view of experts involved in the protocol analysis.

The eight remaining problems that are not used for the sequential problems test are used as a basis for developing items for the enabling skills test.

Step 2 - Using the individual protocol analysis, gain an understanding of the technical details related to each problem.

Step 3 - Adapt each of the four problems to fit the form of the sequential problems test. For this step, use is made of individual protocol analysis and a generic test shell. The test shell is generic in that all possible electronic components and troubleshooting steps available within the testing problem space are included in the shell. To adapt the generic test shell to individual problems, electronic components and troubleshooting steps are eliminated to match the testing problem space for individual problems. The hypothesis list associated with each problem is also individually tailored using the test shell. Extensive use is made of the effective problem space graph for each of the problems. An example is shown in Figure 3. Additional features obtained from the use of a test shell include uniformity of language and option lists across individual problem tests. The skills analysis graph shown in Figure 4 is also used at this step to highlight locations in the problem space where skills of interest can be measured.

Step 4 - Organize the cognitive skills to be assessed in the enabling skills test and prepare initial drafts of the test materials (i.e., situations, test items, related schematics and computer code, scoring key).

Step 5 - Conduct an extensive review of the test materials using subject matter experts. Check material developed in steps 3 and 4 for:

a. Factual correctness

b. Match to the skills they were prepared to measure

c. Correct use of technical language

d. Freedom from bias

e. Appropriateness of branching for individual problems within the sequential problems test

f. Discriminating power (experts should perform at least as well as novices on all testing materials)

g. Consistency with correct item-writing principles

A survey form is developed for each section of the DAT to systematically review content issues.

Step 6 - Revise testing materials based upon the test reviews.

Step 7 - Develop orientation material for both the enabling skills test and the sequential problems test.

Step 8 - Enter the text and branching parameters for the DAT into existing software.

Step 9 - Conduct a pre-pilot of the DAT using a sample of four airmen (2 novices and 2 experts). The purpose of the pre-pilot is to check:

a. computerized administration of the DAT

b. clarity of the test orientation and individual item directions

c. testing completion time

d. performance of high and low performers

e. completeness of option list and necessary technical orders

f. airmen's reactions to the test.

Information is obtained through the questioning of airmen during testing and post-test interviews.

Step 10 - Revise testing materials based upon the results of the pre-pilot.

Step 11 - Prepare materials for pilot administration. Design pilot studies. Choose samples, sites, etc.

Step 12 - Conduct a pilot administration with as large a group of expert and novice airmen as is reasonable, and, in addition to those topics listed for the pre-pilot, check:

a. scoring keys

b. acceptability of item statistics and distractor effectiveness

c. skill reliability

d. test reliability

e. test score validity

Supervisor ratings of airmen's technical skills and probability of completing individual problems within the sequential problem tests are collected prior to the pilot study to help in identifying experts and novices, and providing a database for the subsequent validity investigation. Again, airmen are interviewed during and after testing for comments.

Step 13 - Revise testing materials based upon the results of the pilot administration. Assemble new directions, items, scoring keys, etc.

Step 14 - Design and conduct additional reliability and validity studies.

Step 15 - Prepare a technical manual and final version of the test.

These 15 steps are an update of the steps originally proposed in our work. Based upon three years of experience, the current 15 steps are very different from the original ones (see, Hambleton, 1986).

Special Problems in Test Development

The complexity of the Air Force specialty has implications for the development of the sequential problems test:

1. Determining an appropriate testing problem space

Following up every possible troubleshooting step an airman could take with the appropriate branching and results is prohibitive. Test development and administration time would be greatly increased and, most importantly, the skills that would be assessed for many "unreasonable" steps are not known. The problem space for the sequential problems test consists of novice, mid-level and expert solution paths with follow-up branching and results for only <u>common</u> inappropriate steps. Inappropriate <u>uncommon</u> steps are not usually available from the cognitive task analysis prepared in Task I of the project. To obtain them, would be extremely time consuming, very costly, and of limited value in the test development work anyway.

2. Responding to troubleshooting steps not in the testing problem space

As stated previously, those inappropriate uncommon troubleshooting steps that are available are not followed up with branching or results. The test taker must, in these cases, be told why the step is not followed up. One option is to give a generic message for each particular type of step. For example, an airman who chooses to swap inappropriate uncommon components would receive a message that reads: "A working space is not available at this time." However, messages of this type can lead to confusion since the options which are followed up vary across the four problems in the test. An airman who tries to swap a component in problem one and receives a "not available" message may think that component is not available for swapping in later problems. At the present time, we have opted for a strictly generic "stop" message: "Please try another option." The effect of receiving any message of this kind on an airman's test performance is, of course, a concern.

3. Determining those technical orders that correspond to the testing problem space

Allowing access to all technical orders used in this specialty or most specialties is prohibitive. The DAT testing area could not accommodate the number of volumes, and reproduction costs for a duplicate set are high. For these reasons, airmen taking the sequential problems test have access to technical orders that correspond to novice, and mid-level and expert solution paths. Additionally, technical orders related to common inappropriate steps and "distractor" technical orders are also available during testing.

4. Developing branching that allows ease of movement through the testing problem space

By following up only a subset of the total available troubleshooting steps, the testing problem space has been reduced. However, the testing space is still quite large. An airman may choose from hundreds of possible troubleshooting steps in a typical problem of the sequential problems test.

Two branching schemes were considered: (a) task-controlled and (b) location-controlled. Task-controlled branching begins with a menu of generic troubleshooting steps ("Take a measurement," "Run a programmed test"). Subsequent branching offers menus that more specifically identify the chosen action and present the results of that action. Alternately, location-controlled branching consists of determining in what physical area the airman wants to investigate. The airman is then presented with an appropriate menu of troubleshooting steps for that location. Our present sequential problems test uses task-controlled branching. Cuing is reduced with task-controlled branching since all possible micro and macro level troubleshooting steps are available from the main menu.

Issues in Test Development

The development of the DAT has raised many other interesting psychometric and practical test development issues. A few of these issues will be discussed next.

First, the original format of the sequential problems test segmented each problem according to the level (macro/micro) of troubleshooting. Each segment presented a menu of actions appropriate to the level of problem solving. The airman would repeatedly see and choose from this menu until a decision was made which routed him/her to the next more detailed segment of the problem.

The reasons for originally structuring the sequential problems test according to this format included:

a. The menu of possible actions offered for each segment of the problem could be comprised of those actions and distractors which were appropriate for that segment.

b. The possible problem solution paths were constrained somewhat by our bringing the airmen together at decision points.

c. It could be assumed that at decision points the airmen would all have acquired information from the preceding segment of the problem but would not yet have accessed the more detailed information of the next segment. This assumption allowed us to ask questions at decision points of all airmen and minimize the problem of airmen approaching the questions with variable experience and knowledge.

d. Having the problem segmented offered more structure for our job of scoring and analysis.

After a period of evaluation, a major problem with this format emerged. The troubleshooting behavior of airmen does not necessarily progress linearly from a macro to micro level. Rather, there tends to be considerable movement between levels. Our original format for structuring the sequential test did not allow for this mobility and, therefore, did not properly simulate the problem-solving behavior of airmen. Several small modifications to this format were considered. However, in modifying the format, we found that many of the advantages of the original format were no longer available.

The present sequential problems test makes use of task-controlled branching (described earlier) which allows airmen complete mobility between macro and micro troubleshooting steps. In terms of fidelity and validity, this format seems preferable. The major drawback of this format is that scoring becomes more complex as the testing space is no longer constrained to more well-defined segments.

Secondly, the development of the enabling skills test involved making use of unusual item formats that include master lists, ranking/sequencing and judging relevance. The use of these formats allows for the measurement of skills deemed necessary to expert performance in electronics specialties. For example, airmen must have knowledge of potentially dangerous actions performed during electrical tests; therefore, judging relevancy items that ask an airman if particular troubleshooting steps given a specific situation are dangerous, unnecessary, or necessary, are used.

Fidelity and validity of the enabling skills test may also be increased through the use of job-relevant stimuli. Item stems often include "on the job" troubleshooting situations, computer testing code or lists of test equipment. In addition, airmen are of en asked as part of an item stem to view more extensive technical orders in a test reference book.

Another test development issue is reducing the time required to develop simulation-type tests. The development of the individual problems within the sequential problems tests required many repetitive tasks. Lists of electronic components and possible troubleshooting steps had to be developed and much time was spent on reproducing similar lists for each problem. For this reason, we adapted the use of a generic test shell. The test shell is generic in that all possible electronic components and troubleshooting steps available within the testing problem space were included. To adapt the generic test shell to individual problems, electronic components and troubleshooting steps were eliminated to match the testing problem space for that problem. Additional features obtained from the use of a test shell include uniformity of language and option lists across individual problem tests.

Conclusions

After three years of research on this project, we feel that a reasonable test design is in place, and our steps for constructing DATs appear to work. Also, the computer software is in place and appears to be running correctly. Finally, our initial pilot testing was informative and supportive of the basic testing approaches, but revealing of a number of problem areas such as excessive test length, glitches in the computer software, and flaws in our test orientation and approach to obtaining airmen's hypotheses about the locations of faults as they work through the problems. With many of the problems behind us, we are now ready to begin a series of extensive validity studies to investigate the test design, scoring methods, report forms, and the validity of the cognitive scores produced from the test.

Labrepor.1.30

References

Gott, S. (1988). Assessing technical expertise in today's work environments. In Proceedings of the 1987 ETS Invitational Conference on Testing Problems. Princeton, NJ: Educational Testing Service.

Hambleton, R. K., Gower, C., & Bollwark, J. (in preparation). Construction and scoring of computer-administered diagnostic tests to assess the cognitive skills of airmen in technical specialties. Laboratory of Psychometric and Evaluative Research Report. Amherst, MA: School of Education, University of Massachusetts.

Hambleton, R. K. (1986). Development of an integrated system to assess and enhance basic job skills: Research plan for the personnel measurement subsystem. Air Force Research Report No. 2.. Amherst, MA: School of Education, University of Massachusetts.

Nitko, A. J., & Hsu, T. C. (1984). A comprehensive microcomputer system for classroom testing. Journal of Educational Measurement, 21, 377-390.

Sternberg, R. J. (1985). Cognitive approaches to intelligence. In B. B. Wolman (Ed.), Handbook of intelligence: Theories, measurements, and applications. New York: Wiley.

COGNITIVE SKILLS TAXONOMY

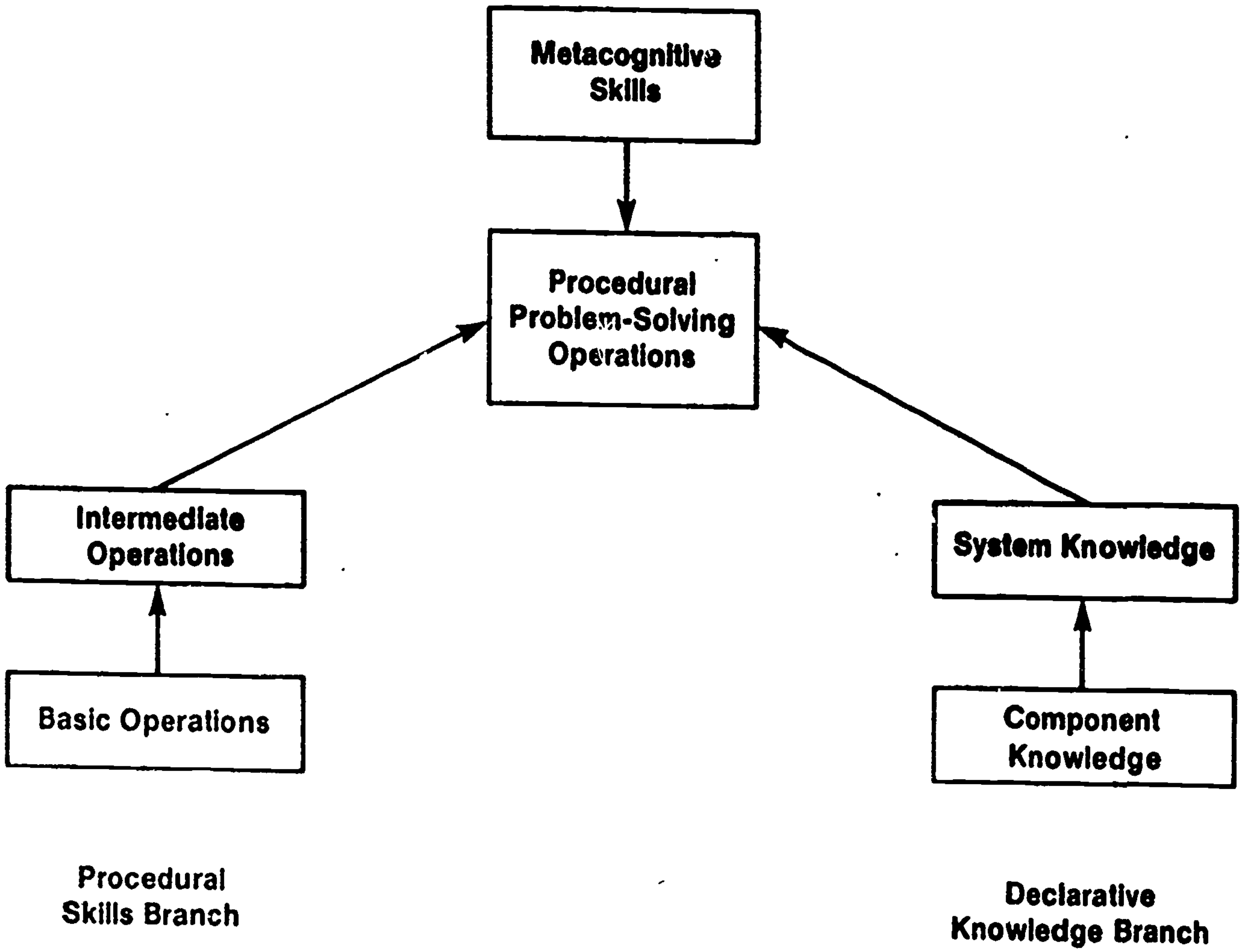


Figure 1. The taxonomy of cognitive skills of interest in the project.

Figure 2. Skills x problems matrix.

SKILLS X PROBLEMS MATRIX 326X4B

				PROBLEMS									
				DATA FLOW		POWER			SIGNAL FLOW				
	SKILLS	Importance	Frequency	Clock Generation	Clock/ Data	Power Supply 1	Power Supply 2	Power Supply 3	Relay	Measur. Input	Sign. Cond.	Wave. Anal.	Bad Pins, Wiring
BASIC OPERATIONS													
Taking Measurements	Use 0-scope to check for signal (present/absent)	H	6	X	X	X			X	X		X	
	Use o-scope to check waveforms (shape)	H	2	X	X								
	Use o-scope to compare onset times of two signals	L	1	X									
	Use DMM to test voltages	H	5			X	X	X	X				X
	Use DMM for continuity checks	H	3			X				X			X
	Use power meter	L	2						X		X		
	Use extender cards	H	3	X	X	X							
	Select test points (on schematics)	H	10	X	X	X	X	X	X	X	X	X	X
	Use logically complete set of test points	H	7	X		X		X	X	X	X		X
	Identify input voltages (using FAPA, TOs, schematics, etc.)	H	2			X							X
Schematic/Block Diagram Reading	Read block diagrams (understand and use)	H	7	X	X		X		X	X	X	X	
	Read schematics	H	5	X	X			X	X				X
	— Identify components within circuit (inverters, grounds, etc.)	H	7	X	X	X	X	X	X	X			
	— Interpret components in functional terms (inverters, grounds, etc.)	H	3	X					X	X			
	— Know direction you're tracing	H	7	X	X		X		X	X	X		X
	— Perceptual tracing skill	H	6	X	X	X	X		X	X	X		X
	— Use schematic legends	M	7	X	X	X	X	X	X				X
	— Use mental model of component to integrate separate schematics	H	3					X	X				X
	— Use schematic zones (memory load issue)	H	7	X	X		X	X	X	X			X
	Read wiring diagrams (both when useful & how to do)	M	2				X						X
	Use mental model of system to direct tracing	H	6	X	X	X	X		X				X
	Compute logic chip input/output by												
	— reading logic gate diagrams	H	1	X									
	— comparing chip to assumed good chip	L											
	— using knowledge of common chip input/output relationships (e.g., hex-to-decimal converters, J-K flip flops, shift load registers)	H	1	X									
Software Interpretation	Knowledge of FAPA conventions (syntax and semantics)	H	6	X	X		X	X	X	X	X		X
	Knowledge of FAPA subroutines (several lines of code e.g., GOTO)	H	2						X	X			
	Using FAPA coding TO (181-3 for ECP)	H	3	X	X				X				
	Use block of FAPA to identify function of a test (what & why)	H	5						X	X	X	X	X
	Identify and select relevant FAPA statements, pertinent to fail (e.g., routing, test points, limits)	H	6				X	X	X	X	X		X
	Use block of FAPA code to form mental model of portion of system used in test	H	3	X	X		X		X	X	X	X	X
	Use tests pre/post fail to constrain hypothesis set	H	5	X	X				X	X		X	
Use of TOs	Knowledge of TO format/content (e.g., schematic indexes)	H	6			X	X	X		X	X		X
	Knowledge of TO numeration system	L											
	Use of TO table of contents, index (includes knowing names used)	M	2	X									X
Other Basic Operations	Convert between number systems (specify; includes knowing when as well as how to convert)	H	2	oct/dec bin/hex	oct/dec bin/hex								
	Chunk analog and logic circuits into functional areas	M											
	Compute possible inputs given an output	H	1						X				
	Trace a bit entering a card through the card (using schematics)	M											
	Trace data flow from the CCDP out to a TRU & back (using block diagrams)	H											
	Use of basic I/o logic (e.g., if swapping "A" fixes fault, fault in "A")	H	10	X	X	X	X	X	X	X	X	X	X

(Continued)

Figure 2. (continued)

	SKILLS	Importance	Frequency	PROBLEMS: DATA FLOW: Clock Generation	Clock/ Data	POWER: Power Supply 1	Power Supply 2	Power Supply 3	SIGNAL FLOW: Relay	Measur. Input	Sign. Cond.	Wave. Anal.	Bad Pins, Wiring
SYSTEM & COMPONENT KNOWLEDGE													
ATS	Physical/Operational knowledge - UUT/ATS level (e.g. how to connect UUT to ATS)	H	5	X			X		X			X	X
	Functional knowledge - UUT/ATS level (e.g., ATS used to check UUT)	H	5	X			X		X			X	X
	Physical/Operational knowledge - ATS level (e.g, location of controls, how to operate)	H	5	X		X			X			X	X
	Functional knowledge ATS level (e.g., knowledge of functional subsystems)	H	5	X		X			X			X	X
	Physical knowledge - TRU level (generic)	L											
	Functional/Operational knowledge - TRU type level (e.g. how a stimulus drawer operates)	H	3	X	X				X				
	Physical knowledge - specific TRU (specify)	M	8	PPPS ECP	NOISEAN Aux B	CCOP I/O	S/C		MSSU S/C	S/C SWDS Freq Cntr	IFSS S/C MSSU OMM		S/C 07-07 07-01
	Functional knowledge - specific TRU (specify)	H	5	ECP					MSSU	S/C SWDS Freq Cntr	IFSS S/C MSSU DMM		S/C
	Operational knowledge - specific TRU (specify)	H	5	ECP CCDP		CCDP			MSSU	Freq Cntr			S/C
	Physical knowledge - card type (specify)	L	3	dev.contr.	dev.contr.			A3A3A12					
	Functional/Operational knowledge - card type (specify)	H	2	dev.contr.	dev.contr.								
	Physical knowledge - specific card (specify: e.g., can use extender boards)	M	2						A1A7				A4
	Functional/Operational knowledge - specific card (specify)	H	2						A1A7				A4
	Physical knowledge - specific component (specify)	L	4					relay		relay	Pwr head RFgen- erators atten- uators switc.		relay
	Functional/Operational knowledge - specific component (specify)	H	4					relay		relay	Pwr head		relay
Symptom Knowledge	Knowledge of symptoms associated with power problems	M	1			X							
	Knowledge of symptoms associated with data flow problems	H	2	X	X								
	Characteristics of faulty components (specify)	H	4				wires			relay	Pwr head switc. filtrs.		relay wires
INTERMEDIATE OPERATIONS													
ECP (F-15 ATS)	Conduct ECP Serial Loop Test	H	3	X	X				X				
	Use ECP to check status (both how to do and interpret)	M	2	X	X								
	Use ECP to single step through FAPA	H	1										X
	Use ECP to program drawer to perform function	H	3	X	X				X				

-34-

(Continued)

Figure 2. (continued)

	SKILLS	Importance	Frequency	PROBLEMS: DATA FLOW: Clock Generation	Clock/ Data	POWER: Power Supply 1	Power Supply 2	Power Supply 3	SIGNAL FLOW: Relay	Measur. Input	Sign. Cond.	Wave. Anal.	Bad Pins, Wiring
PROCEDURAL PROBLEM SOLVING													
Strategies	Gather more symptom information	H	4	X				X	X	X			
	Distinguish between UUT & ATS fails	H	5	X			X		X			X	X
	Treat problem at macro-level as long as possible	H	3	X		X				X			
	Reconfiguration strategy	M	2						X			X	
	Split-half strategy (i.e., picking intermediate test points)	H	10	X	X	X	X	X	X	X	X	X	X
	Knowledge of software limitations (specify; e.g., what gets checked by confidence vs OA/FI	M											
	"Cheap tricks first" strategy	M	3			X				X	X		
	Use available information to rule out hypotheses	H	6	X			X			X	X	X	X
	Use of functional knowledge to constrain hypothesis set	H	1			X							
	Propogate fault (hypothetical reasoning)	H	5	X	X		X		X				X
	Use knowledge of fault probability (specify component)	H	5	CCOP	CCOP				relay		RF		07-01
Problem Representation	Construct problem representation	H	8	X	X	X	X		X	X	X		X
	Modify prob rep based on discrepant results	H	2			X							X
	Refine prob rep based on additional information	H	5	X			X			X	X		X
	Recognize discrepancy between result & expectation	H	8	X	X	X	X	X	X			X	X
METACOGNITIVE SKILLS													
Working Memory Aids	Record results (written)	M	5	X	X		X					X	X
	Organize results (mentally or graphically)	H	5	X	X		X					X	X
Planning	Develop plan	H	8	X	X		X	X	X	X		X	X
	Adhere to plan (until contra-indicated)	H	6	X	X		X	X	X				X
Self-Monitoring	Choose technique/tool appropriate to your strengths	H											
	Use checking routines (specify)	H	6	re-run test	re-run test		re-run test	re-run test		reconf. fault			re-run test

Figure 3. **Effective Problem Space Graph Measurement Input Problem**

Situation: Freq 1 fails Confidence; numbers keep incrementing; printout reads: "Failed 15180 250 ms Run OA/FI."

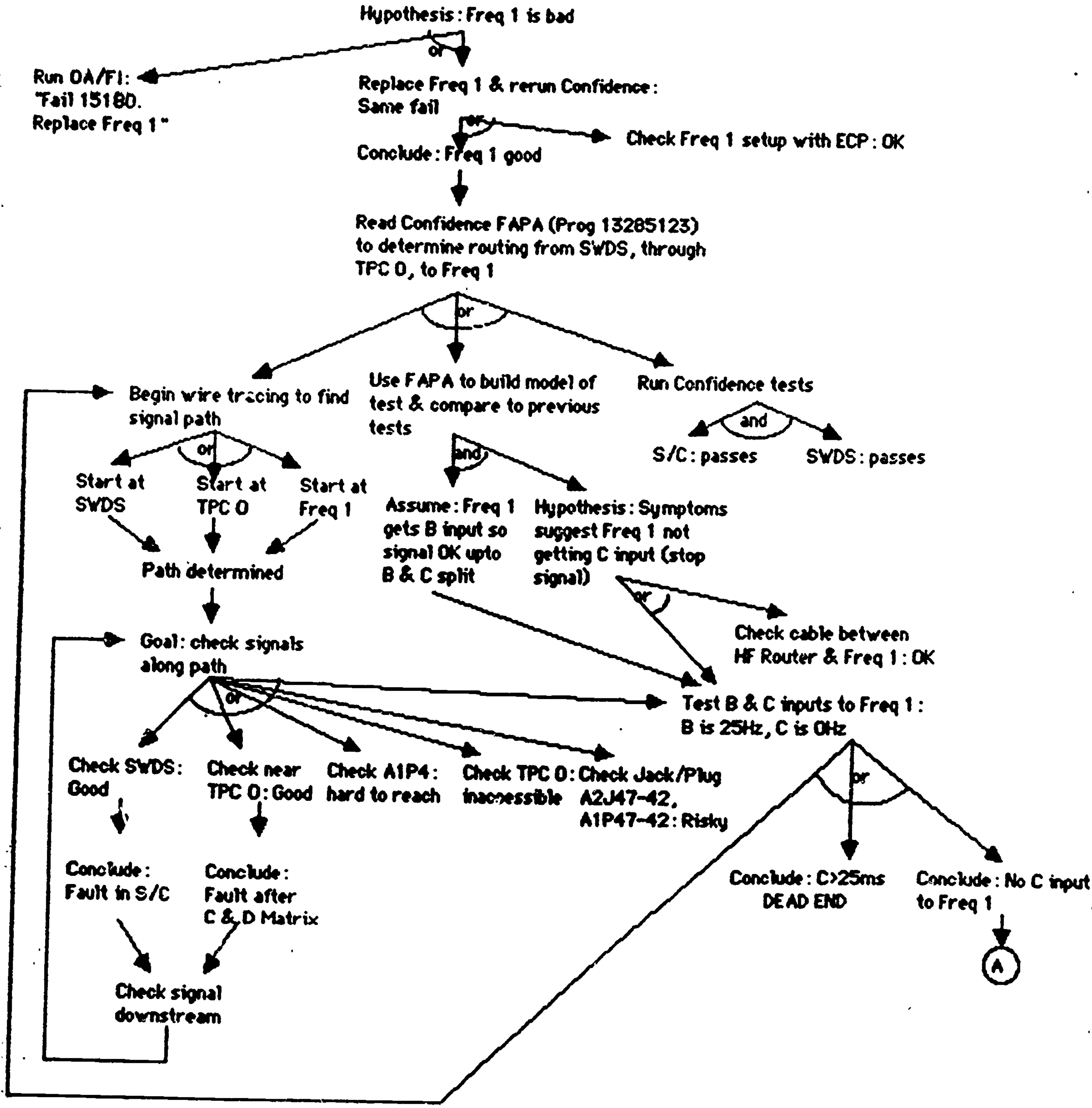


Figure 3. (continued)

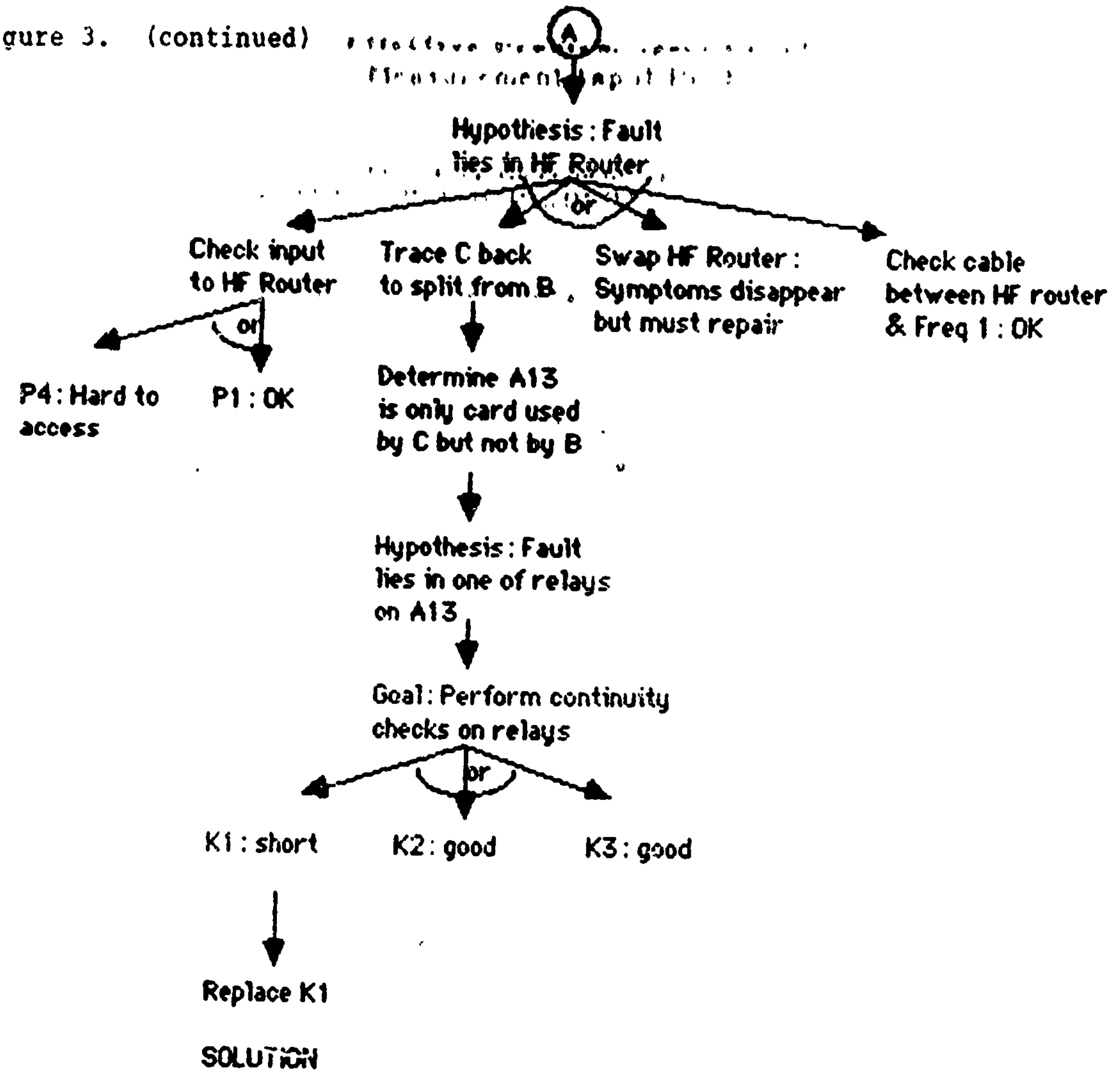
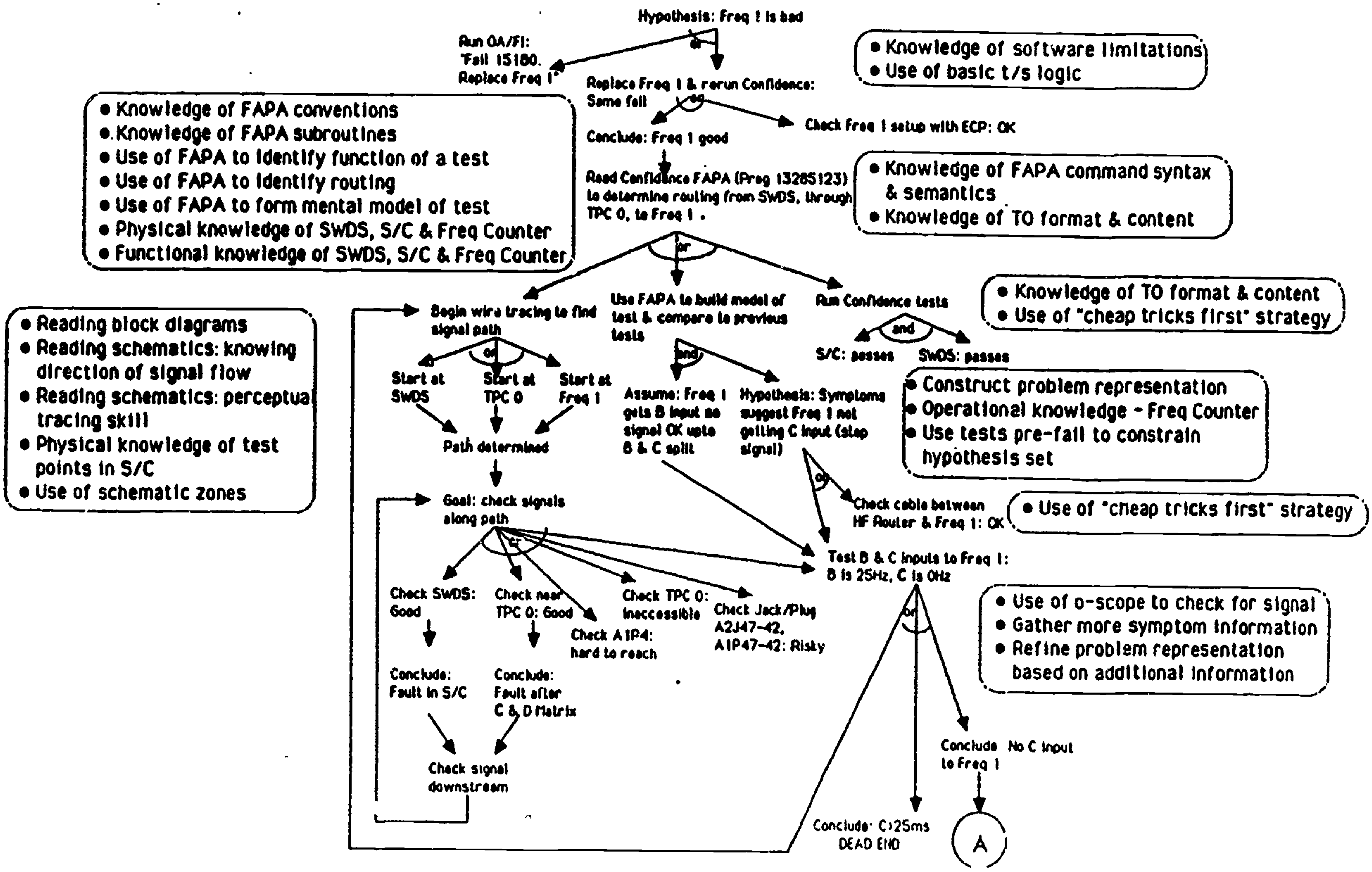


Figure 4. **SKILLS ANALYSIS GRAPH**
MEASUREMENT INPUT PROBLEM

Situation: Freq 1 fails Confidence; numbers keep incrementing;
printout reads: "Failed 15180 250 ms Run OA/FI."



-38-