DOCUMENT RESUME

ED 382 817 CE 068 918

AUTHOR	Gott, Sherrie P.; And Others
TITLE	Tutoring for Transfer of Technical Competence.
INSTITUTION	Armstrong Lab, Brooks AFB, TX. Human Resources Directorate.
REPORT NO	AL/HR-TP-1995-0002
PUB DATE	Feb 95
NOTE	41p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Cognitive Processes; *Computer Simulation; *Electronics; *Experiential Learning; *Intelligent Tutoring Systems; Learning Strategies; Military Training; *Outcomes of Education; Postsecondary Education; Program Evaluation; Technical Education; *Transfer of Training
IDENTIFIERS	Air Force

ABSTRACT

A project was conducted in a military training setting to test the acceleration of the development of complex, technical problem-solving skills through intentional transfer of existing knowledge and skill to novel tasks. Sherlock 2, the model devised, consisted of high quality device models using illustrative equipment diagrams from actual electronic devices; it was used for tutoring the trainees. Coaching on how it works was presented periodically throughout the tutoring sessions. Reflective follow-up included reasoning behind the expert solution and general troubleshooting principles. Partial results of an evaluation of the intelligent tutoring system showed that the tutor's learning environment is concordant with views from the acquired skill through coached apprenticeship activities. Instructional content consists of authentic problem-solving scenarios that situate trainees in realistic contexts where they can practice and hone complex diagnostic skills. Results show that the trainees' acquired expertise is generalizable to a novel equipment system. Similar generalizability was not demonstrated by a matched control group. (Contains 29 references.) (Author/KC)

TM

AL/HR-TP-1995-0002

ARMSTRONG LABORATORY

TUTORING FOR TRANSFER OF TECHNICAL COMPETENCE

Sherrie P. Gott
MSgt Ronald S. Kane

HUMAN RESOURCES DIRECTORATE
MANPOWER AND PERSONNEL RESEARCH DIVISION
7909 Lindbergh Drive
Brooks AFB TX 78235-5352

Alan Lesgold

University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

February 1995

Approved for public release; distribution is unlimited.

BEST COPY AVAILABLE

AIR FORCE MATERIEL COMMAND
BROOKS AIR FORCE BASE, TEXAS

018 990

NOTICES

The Office of Public Affairs has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

Sherrie P. Gott

SHERRIE P. GOTT, Ph.D.
Project Scientist

PATRICK C. KYLLONEN, Ph.D.
Technical Director
Manpower & Personnel Research Division

Gary D. Zank

GARY D. ZANK, Colonel, USAF
Chief
Manpower & Personnel Research Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
February 1995

3. REPORT TYPE AND DATES COVERED
Technical Paper - May 1993 - November 1993

4. TITLE AND SUBTITLE
Tutoring Transfer of Technical Competence

5. FUNDING NUMBERS
PE - 63227F
PR - 2949
TA - 00
WU - 03

6. AUTHOR(S)
Sherrie P. Gott
Alan Lesgold
Ronald S. Kane

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Armstrong Laboratory (AFMC)
Human Resources Directorate
Manpower and Personnel Research Division
7909 Lindbergh Drive
Brooks AFB TX 7823-5352

8. PERFORMING ORGANIZATION REPORT NUMBER
AL/HR-TP-1995-0002

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
To appear in B. Wilson (Ed.), Educational Technology, Special Issue: Designing Constructivist Learning Environments.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

We report partial results of the evaluation of an intelligent tutoring system (Sherlock 2) that fosters the transfer of complex technical skills. The tutor's learning environment is concordant with views from the acquired skill through coached apprenticeship activities. Instructional content consists of authentic problem solving scenarios that situate trainees in realistic contexts where they can practice and hone complex diagnostic skills. Moreover, results show that their acquired expertise is generalizable to a novel equipment system. Similar generalizability was not demonstrated by a matched control group.

14. SUBJECT TERMS
Coached apprenticeship, Cognitive psychology, Cognitive skills transfer, Cognitive task analysis, Constructivist training, Effect size, Evaluation study, Expertise, information bank, Intelligent tutoring system, Intentional transfer, Learning by doing, Phenomenaria, Problem reflection, Simulated work environment, Strategic knowledge, Verbal performance tests, Verbal protocols

15. NUMBER OF PAGES
41

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
Unclassified

18. SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

19. SECURITY CLASSIFICATION OF ABSTRACT
Unclassified

20. LIMITATION OF ABSTRACT
UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

CONTENTS

List of Tables

List of Figures

PREFACE

We would like to thank MSgt Mark Gallaway and Dr Bob Glaser for nurturing the initial inspiration for the study of transfer in technical domains. We are especially indebted to the hundreds of F15 avionics technicians at Langley AFB VA, Eglin AFB FL, Luke AFB AZ, and Nellis AFB NV whose expertise and conscientious dedication to their mission as aircraft maintainers have made this work possible. Thanks is due as well to the HQ TAC/LG (now HQ ACC/LG) staff for their long-term support of the Basic Job Skills (BJS) scientific enterprise.

The quality of the science in this and every other BJS product has been immeasurably enhanced by contributions from our troika of eminent scientific advisors: Dr Robert Glaser (chair), University of Pittsburgh; Dr David Kieras, University of Michigan; and Dr Robert Linn, University of Colorado.

Finally, we dedicate this report and others in the BJS series to Gen Henry Viccellio, Jr., without whose vision and stalwart support this work would have never made the difference it has in the Air Force maintenance community.

The research reported in this paper was conducted in full partnership with the Learning Research and Development Center, University of Pittsburgh, Prof. Alan Lesgold, Principal Investigator. The opinions expressed herein are those of the authors and do not necessarily reflect those of the Air Force. Correspondence concerning this paper should be addressed to Dr Sherrie Gott, Armstrong Laboratory, AL/HRMD, 7909 Lindbergh Drive, Brooks AFB TX 78235-5352.

TUTORING FOR TRANSFER OF TECHNICAL COMPETENCE

COACHED APPRENTICESHIP AS A FORM OF CONSTRUCTIVIST TRAINING

The constructivist movement has grown largely as a reaction to problems in education that come from excessive emphasis on algorithmic performance and rote memorization as the core content of education. It has been noted that saying the words of a principle does not produce understanding of what those words mean. Also, as long ago as Whitehead, it was observed that having learned something in school did not imply that one would use the acquired knowledge when it was relevant in real life. This fundamental problem, that schooling seemed too internally focused and did not prepare people for life outside the context of school performances, led to the constructivist viewpoint that students must construct their own knowledge and that education consists of providing appropriate learning situations that afford a student opportunities to develop personal knowledge that will be useful in later life.

At one extreme, the constructivist approach takes on an idealized humanistic character. Within this view, it is thought inappropriate for the teacher even to have explicit goals for learning. The student is to be free to develop his own mind, his own understanding, his own competence. The viewpoint we have taken is perhaps closer to the other extreme. We realize the need for the student to build knowledge anchored in his/her own prior knowledge and understanding, but we see many situations in which a culture has specific learning goals for its novitiate and in which seeking instruction into that culture implies acceptance of at least a partial goal structure for learning.

Even in this more restrictive view, a constructivist approach is powerful and perhaps necessary. The knowledge a student will acquire must still be anchored in his/her experiences. The nouns and verbs in statements of principle must still have meaning for the student. At the same time, the learning environment must afford opportunities to reflect on how a particular community of practice talks about the world, how it represents the world, and how it determines how to act in the world. We have focused our attention on a particular community of practice, technicians who use and maintain specialized electronic equipment that itself is used to facilitate testing and repair of aircraft navigation equipment. Because this community is defined partly by its responsibilties to a larger community that includes pilots and tacticians who count on having working aircraft, effective practice is constrained partly by cultural responsibilties. It is also constrained by the need to be able to talk to other community members about the work of the community and especially the responsibility to be ready to take on new, but related sets of tasks. Below, we describe and evaluate an approach to training that pursues this particular variant of the constructivist theme.

For the last ten years, we have worked with a team of colleagues from the University of Pittsburgh Learning Research and Development Center and U.S. Air Force Armstrong Laboratory to develop an instructional approach we call intelligent coached apprenticeship[1] (Eggan & Lesgold, in press; Katz & Lesgold, 1991, in press; Katz, Lesgold, Eggan, & Gordin, 1992, in press; Gott, 1987; Gott, 1989; Gott, Hall, Pokorny, Dibble, & Glaser, 1992; Gott, Pokorny, Alley, Kane, & Dibble, in press; Hall, Gott, & Pokorny, in press; Kane, 1993; Lajoie & Lesgold, 1989; Lesgold, in press; Lesgold, Eggan, Katz & Rao, 1992; Glaser, Lesgold, & Gott, 1986; Lesgold, Katz, Greenberg, Hughes & Eggan, in press; Lesgold & Katz, 1992; Lesgold, Lajoie, Bunzo & Eggan, 1992; Nichols, Gott, Alley, & Pokorny, in press; Pokorny & Gott, in press). This approach is based upon the opportunity to experience the most difficult aspects of cognitively-intense jobs in a simulated work environment where assistance, in the form of an intelligent computer-based coach, is always available and where there are opportunities to reflect on simulated work experiences. We have developed two generations of tutors for training a specialized electronics maintenance job in the U. S. Air Force, namely the F-15 manual avionics test station specialty (see Appendix I for an overview of the F-15 Manual Avionics Test Station job). The tutor focuses on the hardest part of the job, isolating failures in the test station itself. Both generations of the training system we have built, named Sherlock 1 and Sherlock 2, have worked remarkably well, in terms of success in fostering high levels of job expertise and, with Sherlock 2, promoting transfer to new electronics troubleshooting tasks on novel equipment.

The primary activity within Sherlock is holistic work, at the highest levels of real-world difficulty, though often this requires coaching, which is available on demand. The approach has several distinguishing characteristics:

- Learning activity is centered in a simulated work environment.

- Learning activity is centered around problems that exemplify the hardest parts of the job for which one is being trained (problems defined in collaboration with master technicians on the job).

For each problem, two kinds of activities occur:

- The student solves the problem, requesting advice from the intelligent tutor/coach as necessary.

---

[1] Sherlock 2, the current embodiment of the ideas we discuss in this paper, has been a collaborative effort that has included Daniel Abeshouse, Marilyn Bunzo, Roberta Catizone, Dennis Collins, Richard Eastman, Gary Eggan, Mark Gallaway, Robert Glaser, Maria Gordin, Sherrie Gott, Linda Greenberg, Ellen Hall, Edward Hughes, Ron Kane, Sandra Katz, Dimitra Keffalonitou, David Kieras, Susanne Lajoie, Alan Lesgold, Robert Linn, Thomas McGinnis, Johanna Moore, Dan Peters, Bob Pokorny, Rudianto Prabowo, Govinda Rao, Rose Rosenfeld, Kurt Strobel, Gary Walker, and Arlene Weiner. Collins, Gallaway, Gott, Hall, Kane, Pokorny, Strobel, and Walker are U.S. Air Force uniformed or civilian employees; David Kieras is at the University of Michigan; Robert Linn is at the University of Colorado; the others are or were at the University of Pittsburgh.

- The student reviews a record of her[2] problem-solving activity, receiving constructive critique from the coach.

Sherlock provides a simulation of the work environment for the F-15 avionics job, using a combination of video and computer graphic displays. Simulated controls can be operated with the computer mouse, and the displays change to reflect an underlying computer simulation of the devices being simulated. Since the fundamental activity of troubleshooting in this job is making tests with meters, this is provided realistically by having icons of meter probes that can be "attached" to video images of device test points.

To complement coached learning by doing, we have developed a collection of tools for post-performance reflection. One provides an intelligent replay of the trainee's actions. A trainee can "walk through" the actions he just performed while solving the problem. In addition, he can access information about what can in principle be known about the system given the actions replayed so far (the work of troubleshooting is mostly the making of electrical measurements and then figuring out which possibilities are ruled out and which supported by the pattern of results). Also, he can ask what an expert might have done in place of any of his actions, get a critique of his action, and have his action evaluated by the system. In addition, extensive conceptual knowledge about the system's functions is available from intelligent hyper graphic displays of an expert's circuit model schematic drawing. In these drawings, the boxes that stand for circuit components are all mouse-sensitive and can "tell about themselves." We have also built a tool for displaying an expert solution to the problem, again with extensive conceptual information available as appropriate to each step. Further, there is an option for side-by-side listing of an expert solution and the trainee's most recent effort.

The tools we have built are motivated by substantial research on the reflective activities that might foster learning. For example, Chi and Van Lehn (1991; Van Lehn, Jones & Chi, 1992) analyzed the activity of more and less effective learners in studying worked-out physics problems. They found that more effective learners showed a different pattern of study, paying more attention to the conditions under which various steps in the solution were taken, to the relations between actions and goals, to the consequences of actions, and to the underlying meanings for formalisms such as equations. Gott, Hall, Pokorny, Dibble, and Glaser (1992) made similar observations in a study where transfer of skill under naturalistic conditions was investigated. Bielaczyc, Pirolli, & Brown (1993, April; Pirolli & Bielaczyc, 1989) demonstrated that students could be taught a similar approach.

Working from the Chi, Bielaczyc, and Gott studies just cited, we can infer several possible roles for post-problem reflection. First, if the trainee reached impasses during his efforts and had to ask for help, then there is some learning work to be done. The trainee must figure out why the

---

[2] To enhance readability, we alternate between masculine and feminine pronouns rather than using more cumbersome forms. About 23% of the target population of the Sherlock system are women.

suggestions of the intelligent coach were useful and what rule(s) can be inferred. Second, problem solving experiences afford opportunities for tuning the generality of procedural knowledge and also for elaborating conceptual knowledge. This is especially the case where intuitive guessing was part of the solution process:

"I tried doing X because it seemed like it might work; why did I think it should work?"

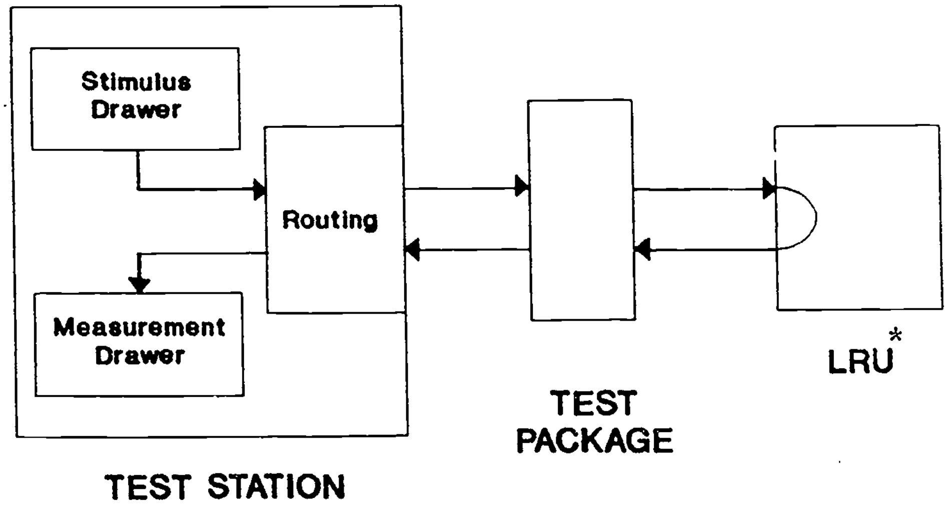
Problems can often be solved in non optimal ways. When this happens, there is no impasse to cue the trainee that his knowledge needs further tuning. So, criticism may be a useful part of the reflection opportunity. Of special relevance are the trade-offs involved in testing hypotheses by swapping parts versus measuring electrical properties of the faulted system. Just as in football, part of what a coach can do is to point out possibilities for improvement that may not be evident to the trainee with respect to cost-benefit trade-offs in the selection of solution steps.

While this instructional approach differs radically from the approaches promoted by traditional instructional design schemes, it is equally dependent upon good task analysis. What is different is that the structure of learning tasks is more authentic, rooted in the needs of practice (or simulated practice) rather than being derived directly from task analysis structure (Gott, 1987; Hall, Gott, & Pokorny, in press).

One important component of our constructivist approach is the intelligent hyperdisplay. When Sherlock constructs a schematic diagram to help illustrate the advice it is providing, that diagram is organized to show expert understanding about the system with which the trainee is working. The structure of the diagram reflects the expert representation of the circuitry involved in carrying out the function that failed, as revealed in a detailed cognitive task analysis. What is displayed is approximately what a trainee would want to know at that time, but every display component is "hot" and can be used as a portal to more detail or explanation. The part of the system on which the expert would be focusing at a given point in the problem solution process is allocated the most space in the diagram and presented in the most detail. All diagram components are "buttons" that can be pushed to expand their level of detail. Boxes in the diagram are color coded to indicate what is known about them given tests carried out so far. Circuit paths are color coded to indicate whether the electrical properties of those paths are known to be appropriate or inappropriate for the function that has failed. Sometimes during problem solving, information is deleted from the display before it is shown, so that the trainee doesn't substitute looking at labels in the displays for inferring what circuitry is involved in the functional failure being diagnosed.

Tutoring for Transfer

Our principal pedagogical goal in initiating the Sherlock project was to <u>accelerate</u> the development of complex, technical problem solving skills. In Sherlock 2, this goal was accompanied by a second, equally important one, namely, to foster technical adaptiveness. Skill flexibility is vital in an era of rapid technological change. Earlier work conducted by our research group in the area of transfer revealed some important findings about intentional transfer, or the undisguised requirement to transfer one's existing knowledge and skill to novel tasks (Gott et al, 1992). Those data revealed that mental models of devices exert strong influence on knowledge access, additional learning (knowledge extension), and subsequent diagnostic reasoning. In the domain of avionics troubleshooting, the primary content of transfer takes the form of abstract knowledge representations. Time and again we observed good learners access their existing mental models of equipment structure and function and their schema of the troubleshooting task (Figures 1 and 2). They then used these models as flexible blueprints to guide their performance as they crafted solutions to new problems. Their prior models became interpretive structures, and when these models were inadequate, better learners flexibly used them as the basis for transposed and elaborated structures that could accommodate the novel situations. They were ready and willing to construct new knowledge that was gounded in their existing representational and functional competence.



*Note: LRU stands for Line Replaceable Unit, a module removed from the aircraft on the flightline.

Figure 1. Top-level mental model of avionics eqipment system

1. RULE OUT UNIT UNDER TEST (UUT) AND TEST PACKAGE

Find out whether problem is in the UUT or the test equipment (i.e., Test Package and Test Station). Note: Once connected to the Test Station, the Line Replaceable Unit (LRU) becomes the unit under test

2. INVESTIGATE MEASUREMENT SIGNAL PATH

Find out the location of the problem on the measurement signal path — either a bad component or one receiving bad data.

3. INVESTIGATE MEASUREMENT DATA PATH

Find the component in the measurement data path that is causing the bad control data input.

4. INVESTIGATE CAUSE OF INCORRECT LIGHT STATUS (ON FRONT PANEL)

Find the component causing the incorrect light state of the indicator which is in the measurement or stimulus data path.

5. INVESTIGATE STIMULUS SIGNAL PATH

Find out the location of the problem on the stimulus signal path.

6. INVESTIGATE STIMULUS DATA PATH

Find the location of the problem on the stimulus data path.

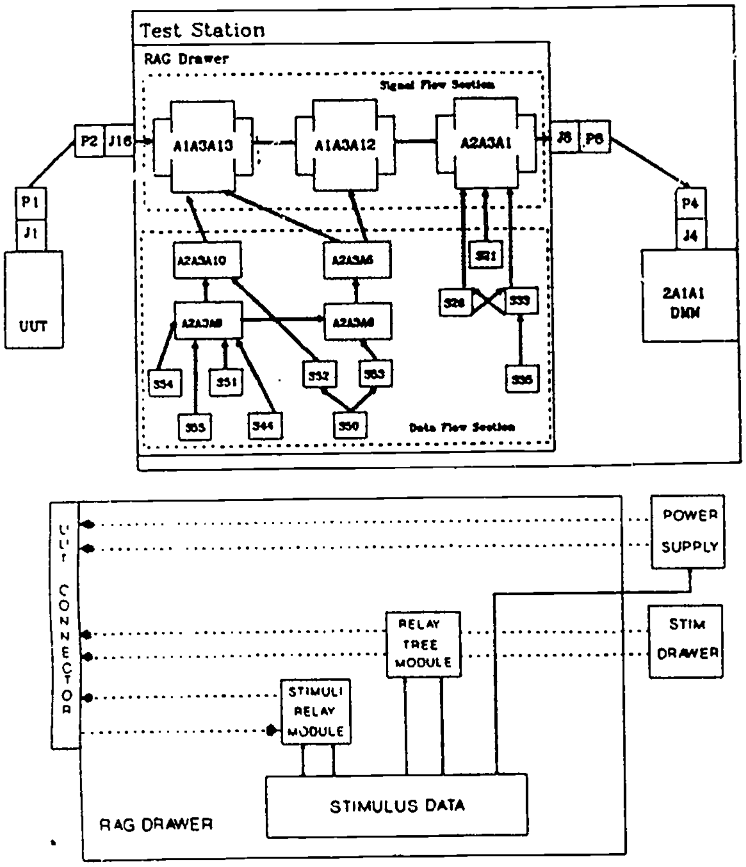


Figure 2. Schema of Troubleshooting Task for F15 Test Stations

By contrast, less able performers devised ways to avoid this adaptive learning experience. They displayed maladaptive behaviors as they oversimplified new problems and overgeneralized existing structures. As a result, their performance in the new domain appeared novice-like, without the benefit of abstract plans and adapted models. They were wedded to their old structures, unable to perceive that the functional variations in the devices in the novel domain were plausible extensions of their current understanding.

These findings in turn influenced instructional design decisions in Sherlock 2, as follows: (a) high quality device models were fostered with liberal scaffolding in the form of illustrative equipment diagrams used in coaching and reflective followup; (b) interactive video representations of actual electronic devices dominated the learning environment (in constructivist terms, the video devices provided phenomenaria to be manipulated and otherwise exercised to build understanding); (c) in the tutor's coaching (or Information Banks) general functional terms were used to describe the electronic devices and tests, not problem specific terminology; (d) general terminology was also be used to characterize the goal structure that typifies "the (generic) plan" for troubleshooting in this domain; and (e) the reasons behind preferred goals and procedures were made explicit to reveal the expert's cost-benefit reasoning in evaluating alternative courses of action.

These design decisions were implemented in Sherlock 2 as follows:

- Equipment diagrams appear frequently in the tutor to illustrate the coaching (an Information Bank) that is available to students as well as the in the reflective followup activities at the end of a troubleshooting session (see Figure 3).

- Functional terminology that reveals the electronic test that is being run is consistently abstract in nature, not specific to the given troubleshooting (scenario) situation being diagnosed (see Figure 4).

- The reasons behind procedural and strategic steps are now made explicit (see Figure 5).

- System or "How It Works" coaching is presented at both general and specific levels of detail (see Figure 6).

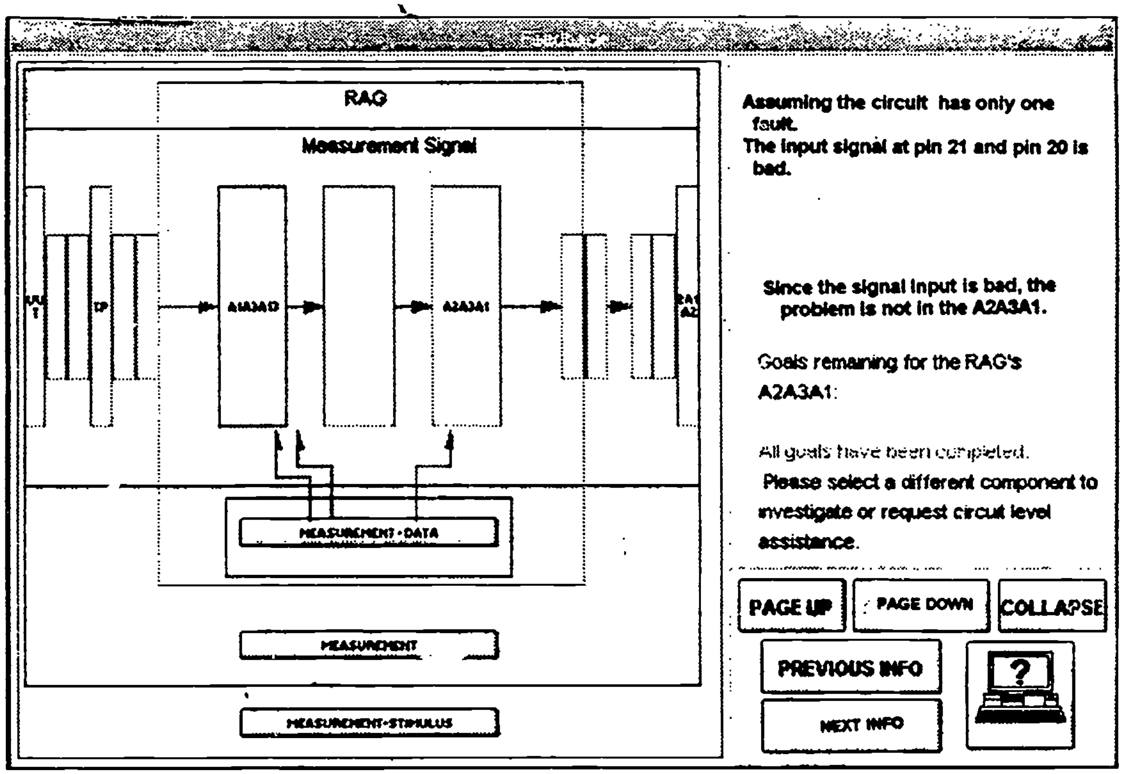
- A standard troubleshooting goal structure is used as a core plan in "How to Test" coaching across all scenarios (see Figure 7).

- In the Reflective Followup, the cost-benefit reasoning behind the illustrative expert solution is made explicit (see Figure 8).

- Also, in the Reflective Followup, general troubleshooting principles are emphasized during the dialog between student and coach (see Table 1).

Table 1. Principles of Good Troubleshooting

1. Perform test procedures safely.
2. Accurately identify <u>active</u> equipment components to investigate.
3. Conduct measurements accurately (correct device, correct signal type, correct probe placement).
4. Test components thoroughly.
5. Swap only after verfying component is faulty.
6. Investigate functional areas of equipment in logical sequence.
7. Make accurate inferences about malfunction based on test results.



Assuming the circuit has only one fault.
The input signal at pin 21 and pin 20 is bad.

Since the signal input is bad, the problem is not in the A2A3A1.

Goals remaining for the RAG's A2A3A1:

All goals have been completed.
Please select a different component to investigate or request circuit level assistance.

Figure 3. Illustration of Equipment Diagram Used in Reflective Followup Activity

CIRCUIT LEVEL COACHING:

There are two main starting points for tracing the measurement signal path: the UUT Test Package (TP) column of the checkout Tech Order, and the switch settings on the RAG drawer.

From the UUT TP column:

1. Look at the UUT TP column to find out which UUT pins are being tested.

2. Examine the test package schematics to identify the test package plug and pins that are connected to these UUT pins.

3. Trace from the active plugs and pins on the test package to the active jack and pins on the RAG drawer.

4. Use the jack/switch index located in front of the RAG schematics to find the zone in the RAG schematics where these active RAG jack and pins are located.

5. Go to this zone, and trace the rest of the measurement signal path from the active RAG jack and pins to the measurement device.

From the Switch Settings on the RAG Drawer:

1. In the checkout Tech Order look at the settings...

•
•
•

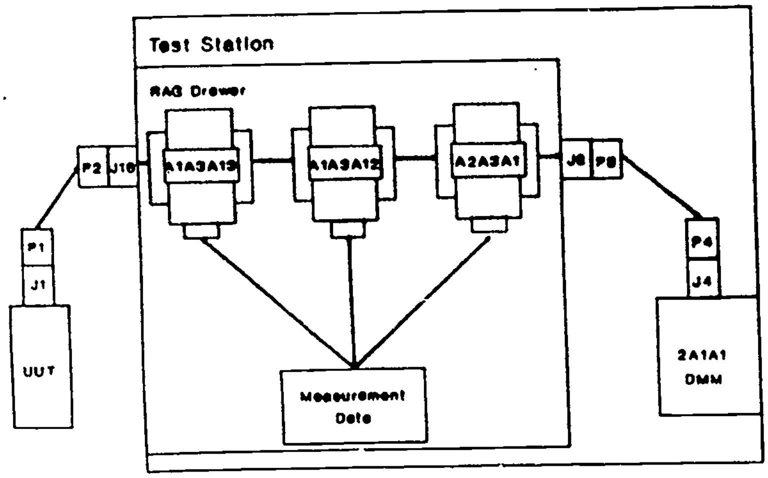


Figure 4. General terminology used in coaching regarding electronic test being run

Step from a Master's solution trace as annotated in the Reflective Followup:

GOAL: To find out whether the problem is in the UUT or the test equipment:

PROCEDURES: The Master technician:

- replaced the UUT's A4 card; reran test; still failed
- made the following test at Jack 1 (J1) of the UUT: test Kohms from J1 pin 1 to J1 pin 2. The reading was 24.4639 Kohms. This was the expected result.

Figure 5. Reasons Behind Procedural Step Made Explicit

TEST POINT SELECTION RELAY CARD - HOW IT WORKS

Level 1: The RAG's A1A3A13 card is a relay card. In a relay card, data signals are applied to the coils of the relays on the card in order to set (or reset) the relays. In this way, one or more paths are created through the card so that signals can travel from the card input(s) to the output(s).

Level 2: The RAG's A1A3A13 is a Test Point Selection relay card. A Test Point Selection relay card consists of ten relays and all ten relays share a common output but have different inputs. In this way, a Test Point Selection card is designed to create only one path between its output and one of its ten input pin pairs. This means that one Test Point Selection card is capable of creating ten different relay circuit paths, but should have only one relay active for any one T.O. test.

Level 3: To work properly, the RAG's A1A3A13 should have only one of its ten relays selected at any one time.

For this test step, Switch 53 is set to 0, Switch 52 is set to 0 and Switch 54 is set to B. This means that 28VDC should be present on pin 33, 0VDC on pin 30, and relay B00 should be set.

When relay B00 sets, a path is created between input pins 51 and 52 and output pins 11 and 12.

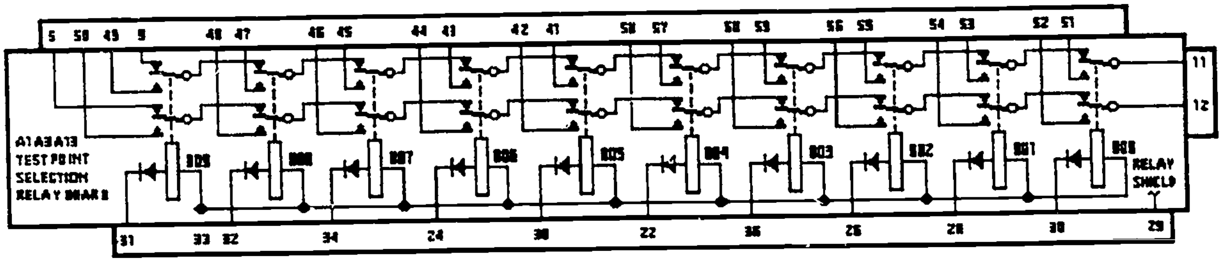


Figure 6. "How It Works" coaching at general and specific levels.

1. Rule out UUT and Test Package first.
2. Investigate measurement signal path next.
3. Investigate measurement data path next.
4. Find source of incorrect light state.
5. Investigate stimulus signal path.
6. Investigate stimulus data path.

Figure 7. Standard Goal Structure Used in Coaching

Illustrative Explanation from a Master:

After verifying that the UUT and Test Package are good, you now know that the problem is in the Test Station. The measurement signal path in the station, i.e., the path between the UUT and the station's Digital Multimeter (DMM), is a good place to start. Testing the signal path allows you to investigate a smaller and somewhat easier functional area of the station, and so you can constrain your search in an effective way.

If you bypass the measurement signal path and start in the measurement **data** functional area, you have a larger area with more difficult components to troubleshoot. The measurement data functional area includes logic cards and switches, for example. The components that send control voltages to relay cards on the measurement signal path are among the components located in the measurement data functional area.

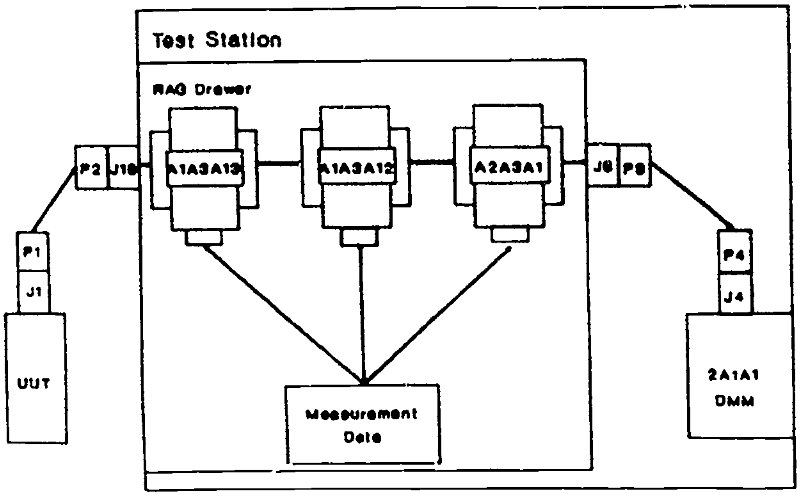


Figure 8. Cost-Benefit Reasoning Behind Master Solution Step

Purpose and Rationale for Tutor Evaluation Study

With these major instructional design decisions having been implemented, the goal of the Sherlock 2 evaluation study was to determine if an intelligent tutoring system that is informed by detailed cognitive models of troubleshooting performances is effective in both accelerating skill acquisition and fostering adaptive expertise. A controlled experiment was conducted that involved U.S. Air Force technicians at three geographically separated F15 flying wings.

Trainees were evaluated on a number of technical proficiency and experience indicators to establish matched experimental and control groups. The principal form of assessing troubleshooting skill was a type of structured, thinking aloud verbal protocol, called a Verbal Troubleshooting Test (VTT). The advanced, most experienced technicians in the shops were also tested on VTTs and other assessment instruments in order to establish the upper limits of performance for this job.

It was hypothesized that the experimental group would demonstrate an accelerated rate of skill acquisition compared to the control group trainees in moving toward the level of performance displayed by advanced technicians. These expectations were predicated on the following premise: in a learning environment that provides direct but coached problem solving experiences and one in which cognitive skill components and processes have been precisely identified as instructional targets, the acquisition of complex skills such as electronic troubleshooting can be speeded up. The expected accelerated acquisition would be attributable to (1) the better instructional content enabled by cognitive models that make the unobservable facets of troubleshooting knowable by learners and (2) better methods where knowledge is tied to its uses in the world and learning is supported by direct manipulation experiences (phenomenaria), coaching, modeling, and other scaffolding embedded in the computer tutor.

It was further hypothesized that the experimental group would demonstrate adaptiveness in their newly acquired troubleshooting skills compared to the control group when tested on a novel equipment system. These expectations were predicated on several premises. First, a learning environment where extensive practice is available to trainees would build up robust task schemas supported by conceptual support knowledge that explains the "reasons why" tasks are structured the way they are. Thus, knowledge structures would be both robust and flexible. Secondly, all coaching and post-session reflective feedback would provide general as well as task specific explanations to inject elasticity into system, procedural, and strategic knowledge components.

METHOD

Subjects

Forty-one Air Force apprentice and thirteen master technicians, for a total of 54 F15 avionics maintenance technicians, participated in the evaluation study. Assignment to either the apprentice or master group was based on VTT proficiency scores as well as on other pertinent factors, such as aptitude and experience. Technicians in this specialty work in a shop environment, vs. on the flightline. Their most challenging tasks involve troubleshooting the test stations that simulate the jet for purposes of testing components (black boxes) that are removed from the aircract (by flightline technicians) because they are presumed faulty. Subjects were selected from three geographically separated sites -- Langley AFB VA, Nellis AFB NV, and Eglin AFB FL. Table 2 shows the distribution of subjects across the three sites. The apprentice technicians had an average of 33 months experience on the job, while the master technicians averaged 124 months (10 years 4 months) experience.

Table 2. Number of Subjects by Groups Across Sites

	Sites		
Group	Langley AFB VA	Nellis AFB NV	Eglin AFB FL
Experimental	6	12	0
Control	4	1	18
Master	2	4	7
TOTAL	12	17	25

Instruments

Four different instruments were used in the Sherlock 2 evaluation. Two were assessments of learning: Verbal Troubleshooting Tests (VTTs) and Noninteractive Troubleshooting Tests (NITs); one was a Background Questionnaire to document the subjects' personal data; and one was a Tutor Report Card, which tutored airmen used to evaluate various dimensions of the tutor as an instructional system. We will only treat the VTTs and NITs here. See Gott, Pokorny, Alley, Kane, and Dibble (in press) for a complete description of all instruments.

Verbal Troubleshooting Tests (VTTs)

An accepted principle in practice-oriented (vs. academic) instruction is that both training and testing should mirror the criterion performance (Gott, 1987). This principle applies to modern cognitive apprenticeships as well; however, the mirroring process is complicated by the fact that performances in such apprenticeships are more mental than physical. No longer is it effective to focus on (diminished) overt behaviors and observable end products for either instructional or assessment purposes. Internalized cognitive processes and structures are the legitimate targets of assessment. For these reasons, we developed special purpose instruments to evaluate the Sherlock tutors. To access and measure the covert processes and structures that Sherlock targets (and that, we would argue, really explain competence), we developed Verbal Troubleshooting Tests (VTTs) as the principal measure to evaluate tutor effectiveness (Gott, 1987; Gott et al, in press). The VTTs satisfy the principle of mirroring the criterion performance and can accommodate the assessment of internalized cognitive processes and structures. The general form of the VTT closely resembles the structured thinking aloud verbal protocol process used in conducting the cognitive task analysis for Sherlock 2 (Gott et al, in press) (see Figure 9). The diagram in the figure is drawn by the subject to illustrate the solution step and thereby reveal the mental model he is using.

PRECURSOR

I want to see if the LRU ID resistor is good.

ACTION

Remove the cable from J12 of the LRU and ohm out the path through the LRU from pin 88 to pin 128.

RESULT

The reading is 1.55 Mohms.

INTERPRETATION

The problem isn't in the LRU, it's in the test station or the test package.

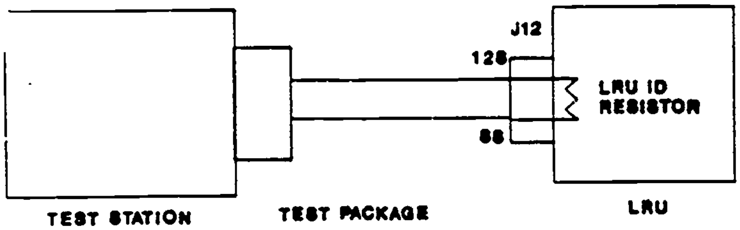


Figure 9. One Solution Step Illustrating the PARI Cognitive Task Analysis Methodology

The basis for the VTT is a structured example of a verbal protocol technique. It is structured in the sense that subjects are queried according to a standard framework in addition to being asked to think aloud in general. The imposed structure reflects the recursive nature of diagnostic reasoning: hypothesize, test (collect data), interpret data, hypothesize, test, interpret, and so forth, until a solution state is reached.

The most valued feature of the VTT is that it allows a close approximation of hands-on performance, without the costly and inefficient utilization of actual equipment. In the Sherlock project, the VTTs were designed to represent the domain by covering critical areas identified by the cognitive task analysis. The CTA results produced a functional model of the equipment configuration that was shared by expert troubleshooters (refer to Figure 1). We in turn have considered this equipment model and the actions used to investigate it to be "the domain" to be covered instructionally. The basic functionality works as follows: in a simulation of in-flight operation, the stimulus generators inside the test station produce signals that are routed via the Relay Assembly Group (RAG) drawer, through the test package, to the UUT. Signals returning from the UUT pass through the test package and RAG drawer on their way to a measurement device in the test station. The technician monitors the outputs of the measurement devices and compares the actual readings with the expected readings listed in a check-out tech manual. The manual lists the expected values as well as the probable source of fail for each test that is run on the UUT.

Avionics experts with considerable test station troubleshooting experience used the CTA results and Sherlock 1 findings to select or develop the VTTs for Sherlock 2. There were two pretest VTTs and two posttest VTTs. During data analysis, one pretest VTT had to be eliminated because it covered content in the tutor that was deleted from the treatment because of time constraints in the field.

Administration. Each pretest and posttest VTT was individually administered by an experienced F15 avionics expert from our research team. The examiner instructed each subject to troubleshoot in the same way as she would in the actual work environment on the real equipment. In each VTT, the examiner posed a situation to the technician where a fault had occurred during the routine benchchecking of an LRU. The situation for each VTT corresponds to a Problem Statement similar to that depicted in Table 3. The examiner begins by asking, "What is the first thing you would do in the shop if this happened?" and the technician responds with an action, thereby beginning the recursive diagnostic reasoning process. The step-by-step solution process engaged in by the subject is very similar to troubleshooting the actual avionics equipment. A series of recursive action-result steps unfolds and constitutes the solution (Steps 1-7 in Table 4.) Ninety minutes were allowed for completion of each VTT.

Table 3. Exemplar VTT Problem Statement

While running a Video Control Panel Unit, Test Step 3e fails. The panel lamps do not illuminate. All previous test steps have passed.

Table 4. Exemplar VTT Solution Trace

Problem statement: While running a Video Control Panel unit, Test Step 3e fails. The panel lamps do not illuminate. All previous steps have passed.

Step 1: What would you do first if this happened in the shop? **Result:** Test still fails	**Action:** Reenter Step 3e instructions
Step 2: What would you do next? **Result:** Test still fails	**Action:** Swap UUT with shop standard; rerun test
Step 3: What would you do next? **Result:** Test still fails	**Action:** Swap in a known good test package; rerun test
Step 4: What would you do next? **Result:** 0 volts AC	**Action:** Take an AC voltage test at output of Manual Stimulus Assembly
Step 5: What would you do next? **Result:** 28 Volts DC	**Action:** Take a DC voltage test at data flow into Card A22
Step 6: What would you do next? **Result:** 1999K ohms	**Action:** Make ohms test from relay card input to output
Step 7: What would you do next? **Result:** Test passes	**Action:** Swap card A22; rerun test

Scoring. To handle the scoring demands of nonstandard data such as VTT protocols, we developed an objective scoring system for Sherlock 1 using policy capturing techniques on judgments which were generated by experts as they evaluated VTT solution traces (Gott, 1987; Pokorny & Gott, in press). From this data base of explicated judgments from multiple experts, it was possible to extract a scoring policy that can be defined by seven principles of good troubleshooting (see Table 1). These principles provide the foundation for the detailed scoring rules that constitute the complete scoring system (Gott et al, in press). These general principles figure prominently in the assessment system used in Sherlock 2.

For Sherlock 2, the interrater reliabilities (based on holistic VTT scores) for the three tests were .96, .95, and .94. For all three correlations, the 1-tailed significance level was p <.001.

Accompanying materials. In accordance with our goal of ensuring that Sherlock is a natural extension of the technician's real work environment, there are two standard pieces of technical material that are needed to work the VTTs as well as perform maintenance tasks on Sherlock: LRU Check-out Procedures and Test Package/Test Station schematic diagrams. The Sherlock analog documents closely mirror the real world tech data. There is also a User's Manual to assist the technician in operating the computer tutor.

Noninteractive Troubleshooting Test

The Noninteractive Troubleshooting (NIT) test (the second instrument to be discussed here) is a paper-and-pencil instrument designed to complement the VTT as a measure of troubleshooting skill. It was developed for two reasons: (1) to provide a basis for a convergent measure of troubleshooting proficiency (with the VTTs) and (2) to examine alternative hypotheses (aside from the Sherlock intervention) that could plausibly explain the expected pre- to posttest VTT gains. This test differs from the VTT in several ways: first, it is a written test administered under highly standardized conditions with a limited role for the examiner, and second, questions require short, unambiguous, written responses, even though the realistic troubleshooting context associated with the VTT is preserved as much as possible. Parallel forms were developed for the NIT Pretest and the NIT Posttest. For untutored subjects, pre and posttest NIT scores were significantly correlated (r = .76, one-tailed significance p = <.001).

Other Instrumentation and Materials

The Novel Equipment System

To investigate the generality of troubleshooting skills acquired by subjects tutored on Sherlock 2, an avionics expert on our research team created a mythical equipment system called Frankenstation (Kane, 1993). Frankenstation is an automatic test station, not the type of manual station represented in Sherlock. Its function is to test components of the Integrated Ground Operational Radar (IGOR) system. To achieve that purpose, it has the same functional areas that are common to any avionics test station (see Figure 10). Its primary uniqueness (vis-a-vis Sherlock) is that it is automated, meaning it is computer (not human) controlled. Therefore, Frankenstation tech data includes programming information for the ATLAST programming language, for example, which would be novel to manual test station maintenance personnel.

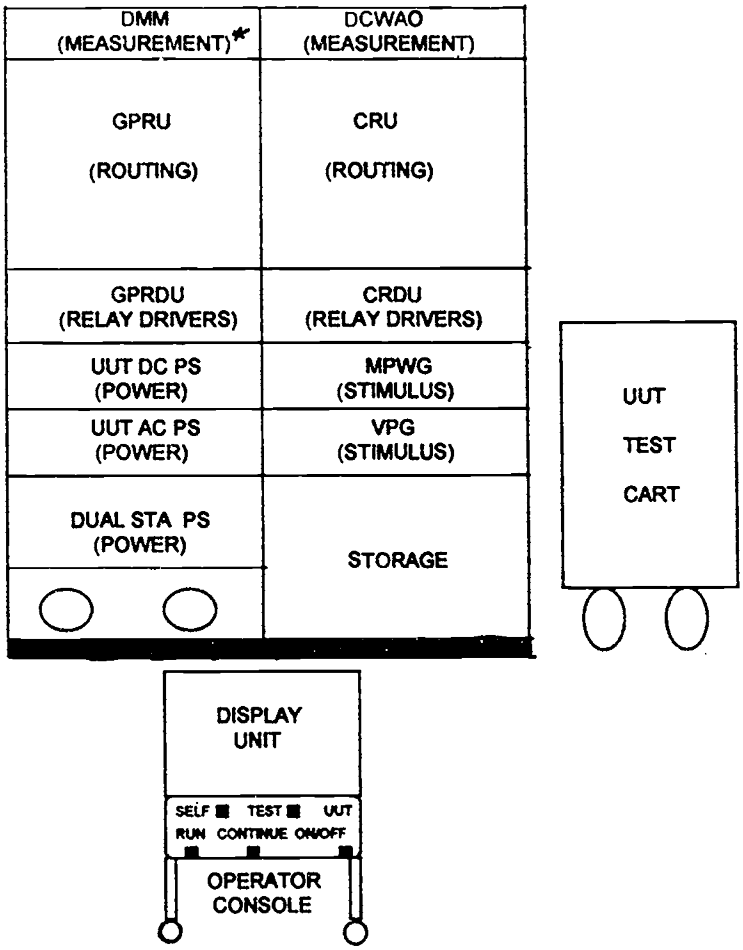


Figure 10. Top-level Functional Diagram of Frankenstation

Note: *Functional terms such as stimulus, measurement, and routing were not provided to subjects on this diagram. They had to consult other technical data to make those functional determinations

RESULTS

Comparability of Groups Prior to Intervention

Initial comparisons of the experimental and control groups on the VTT and NIT pretests and on other performance-related measures (aptitude and experience) revealed no statistically reliable differences prior to the intervention (Table 5). This indicates that the groups are equivalent for purposes of this evaluation study.

Table 5. Pretest Measures of Aptitude, Experience, and Troubleshooting Proficiency

		Aptitude Measures		Experience	Troubleshooting Proficiency	
Group	N	Electronics	AFQT	Manmonths	VTT	NIT
Novices						
Control	23					
M		81	75	27	57	65
SD		10	16	25	36	12
Expermntl	18					
M		84	71	39	39	64
SD		10	14	33	66	20
Masters	13					
M		87	81	124	89	79
SD		11	14	31	9	8

Comparisons of the Novice groups with the Master group revealed expected differences in experience ($t = -9.76$, $p = 0.000$) and troubleshooting proficiency (VTT Pretest: $t = -2.81$, $p = .007$; NIT Pretest: $t = -3.03$, $p = 0.004$). There were no statistically reliable differences (Novices vs. Masters) on the two aptitude measures.

The relationships of VTT and NIT Pretests with the other performance-related measures (aptitude and experience) are shown in Table 6. The strongest correlations were between (1) the two aptitude measures (Electronics and AFQT), (2) one aptitude measure (Electronics) and the NIT Pretest, and (3) the two troubleshooting pretests (VTT and NIT). The only non-significant correlation was between aptitude (AFQT) and experience.

Table 6. Intercorrelations of Pretest Measures

Measure	Aptitude	Experience	Proficiency	
	AFQT	Manmonths	VTT	NIT
Aptitude (ASVAB Elec)	.60**	.32*	.43**	.49**
Aptitude (AFQT)		.28	.41**	.36*
Experience (Manmonths)			.40*	.42**
VTT Pretest				.46**

*p <.01   **p <.001

Post Intervention Results

Experimental and Control Group Comparisons

As predicted, VTT and NIT posttest scores revealed large and statistically significant differences in favor of the experimental group over the controls (Table 7): VTT 3 ($t = -4.04$, $p = 0.000$); VTT 4 ($t = -3.72$, $p = 0.001$); NIT ($t = -2.77$, $p = 0.009$).

Table 7. Posttest Measures of Troubleshooting Proficiency (Sherlock 2)

Group	N	Sherlock 2 Posttests VTT 3	 VTT 4	 NIT
Novices				
Control	23			
M		59	58	75
SD		37	37	14
Experimental	18			
M		95	91	87
SD		5	7	12
Masters	13			
M		85	86	86
SD		12	11	11

The single holistic indicator of VTT performance can be decomposed by an analysis of the components of troubleshooting that are embedded in verbal protocols and known to be associated with expertise, per the cognitive task analysis. Such components include measuring to investigate the equipment rather than swapping equipment parts, systematicity in investigating circuitry versus random actions, and efficiency in targeting areas to investigate by optimizing the information value-to-cost ratio associated with each potential action. Results of the componential analysis are reported in Gott, et al (in press).

Also, as predicted, the Frankenstation VTT and NIT posttest scores revealed large and statistically significant differences in favor of the experimental group over the controls (Table 8): VTT (t = -2.93, p = 0.006); NIT (t = -2.34, p = .025).

Table 8. Posttest Measures of Transfer (Frankenstation)

Group	N	Frankenstation Posttests	
		VTT	NIT
Novices			
Control	21		
M		55	72
SD		31	4
Experimental	17	82	80
M		23	10
SD			
Masters	12		
M		91	88
SD		22	12

The effect size for each of the posttest measures is shown in Table 9. As a basis for comparison, the average effect size for new science and math curriculum in U.S. schools is reported to be .3 sigma (Bloom, 1984).

Table 9. Effect Size for Posttest Measures

Measure	Group						Effect Size
	Control			Experimental			
	N	M	SD	N	M	SD	
VTT 3 (Sherlock)	23	59	37	18	95	5	1.27 SDS
VTT 4 (Sherlock)	23	58	37	18	91	7	1.17 SDS
NIT (Sherlock)	23	75	14	18	87	12	.87 SDS
VTT (Frank'tn)	21	55	31	17	82	23	.96 SDS
NIT (Frank'tn)	21	72	11	17	80	10	.76 SDS

Qualitative Analyses of Frankenstation VTT Data

The single holistic indicator of Frankenstation VTT performance can be decomposed into discrete components of troubleshooting skill that are embedded in the verbal protocols and known to be associated with expertise, per cognitive task analysis data. Such components include (1) thoroughly measuring suspect equipment components during troubleshooting rather than swapping equipment parts prematurely, (2) using automated diagnostic aids (such as system self-tests) in a targeted, efficient manner, (3) following a logical, efficient strategy in troubleshooting, and finally, (4) isolating the fault and thereby solving the problem within the time allowed.

Table 10 shows the comparative swapping actions of the Master, Experimental, and Control Groups. Instances of swapping without testing clearly differentiated the untutored airmen (Control Group) from the Tutored and Master technicians. Table 11 shows how the three groups used self-tests (ST) (diagnostic software) in the course of troubleshooting. While the three groups used the self-test capability at the same relative frequency, the Control Group clearly differed in its manner of implementation. Of the 16 occasions when control subjects used self-tests, there were 11 times when they used the ST inefficiently. They chose to run the ST from the beginning until a fail was encountered as opposed to selecting self-tests that targeted the test station devices that were being used in the failed test. Conversely, Master and Tutored airmen used the ST selectively because they could identify from the program listing and the test station schematics which test station components were being used during the failed test.

Table 10. Frequency of Component Swapping Without Complete Testing (Frankenstation VTT)

	Swaps	Swaps
Group	No Testing	Partial Testing
Controls	50	12
Experimentals	3	13
Masters	8	15

Table 11. Frequency and Quality of Self-Test Use (Frankenstation)

Group	Self Test Use	Frequency of Inefficient Use
Controls	16	11
Experimentals	12	4
Masters	14	2

Instances of following the logical, most effective strategy in troubleshooting Frankenstation clearly favored the Tutored airmen (Table 12). And finally, the Tutored and Master subjects bettered the Controls in terms of isolating the faulty component in Frankenstation i.e., solving the problem, in the time allotted (Table 13).

Table 12. Number of Violations in Logical Sequence of Troubleshooting (Frankenstation)

Group	Number of Violations
Controls	21
Experimentals	9
Masters	20

Table 13. Percentage of Solutions to Frankenstation VTT by Group

Group	Percentage Achieving Solution
Controls	63.6
Experimentals	70.6
Masters	83.3

Discussion of Frankenstation Results

There are a number of intriguing perspectives to adopt in considering the transfer effects from Sherlock 2 to Frankenstation; however, space limitations permit us to explore briefly only three types of empirically supported effects, all of which are related to the findings in Tables 10-13. When we examined the Frankenstation VTT protocols qualitatively, evidence of the transfer of system, procedural, and strategic knowledge was clearly present.

Transfer of System and Strategic Knowledge

The use of self-tests and excessive swapping reveals the differing device models which underpin the actions of Tutored and Untutored airmen. Consider the following solution steps generated by an Experimental and Control subject after each had eliminated the UUT and Test Package from suspicion in Steps 1 and 2:

Experimental	Control
Step 3: Run self-test on the stimulus device. Result: passes	Step 3: Run self-test from the beginning. Result: control/routing device fails; calls out 3 relays as possible causes
Step 4: Run self-test on the measurement device. Result: passes	Step 4: Replace all 3 relays; rerun self-test Result: same fail
Step 5: Run self-test on the control/routing device Result: fails	
Steps 6, 7, & 8: Isolate fail within the control/routing device by measuring at test points that will isolate malfunctioning relay; swap malfunctioning relay; rerun self-test Result: passes	

These data suggest that the Experimental subject is working from the standard device model emphasized in Sherlock 2 (see Figure 1). He looks for and locates devices in the novel equipment that serve the stimulus, measurement, and control/routing functionalities, in effect instantiating the general schema represented in Figure 1. The Untutored airman, by comparison takes an inefficient, "default-type" step that could take hours, i.e., she runs the self-test from the beginning, with no particular device targets in mind. Using the results of the wholesale self-test, she then swaps the three relays, only to learn that the callouts from the self-test do not fix the problem (not an uncommon outcome).

By comparison, the Experimental subject takes measurements, which provide much more information than swapping, and efficiently localizes the malfunctioning relay before swapping any components. The Tutored airman appears to have generalized device <u>and</u> strategic knowledge from Sherlock. This comparison is not an isolated example, but rather the prototypical actions taken by Untutored subjects, as reported in Tables 10 and 11.

Transfer of Procedural Knowledge

Another qualitative inspection of two solution traces from a Tutored and Untutored subject reveals both a specific procedure and a global principle that tutored airmen appear to have generalized from Sherlock 2.

<u>Experimental</u>

Step 5: Rerun UUT; measure ohms from pin to pin on suspect relay.
<u>Result:</u> 0 ohms

Step 6: Measure ohms from pin to pin on different relay.
<u>Result:</u> open

Step 7: Measure control voltage inputs to faulty relay identified in Step 6
<u>Result:</u> 5 vpp

Steps 8 & 9: Measure remaining control voltages to suspect relay.

<u>Control</u>

Step 4: Run self-test from beginning
<u>Result:</u> Control/routing device fails

Step 5: Measure ohms from pin to pin on one of suspect relays called out by self-test.
<u>Result:</u> open

Step 6: Measure ohms from pin to pin on another suspect relay.
<u>Result:</u> 0 ohms

Step 7: Measure ohms from pin to pin on another suspect relay.
<u>Result:</u> 0 ohms

Step 8: Replace relay tested in Step 5; rerun self-test
<u>Result:</u> same fail

The specific procedure displayed by the Experimental subject in this example is to investigate the control voltages to a relay before swapping it. The global principle is to thoroughly test all possible influences on a suspect component before swapping it. With the Control subject, we see that he was able to identify the faulty relay by taking measurements, not by swapping, as most Untutored subjects did. However, he failed to measure the control voltages before swapping the suspect component. A lack of understanding of the global principle that eschews premature swapping appears to be weak or nonexistent with the Untutored airman.

CONCLUSION

What we have reported here are results from a culminating study in a body of empirical work that has spanned ten years. The instructional approach calls for detailed cognitive task analysis results to enable a learning environment where students construct understanding in authentic contexts, rooted in the needs of practice. Results from other studies of the acquisition of complex, practical skills (such as electronic troubleshooting) have demonstrated how cognitive theories of learning and performance can enable improved apprenticeship training. Improvements seem attributable to two general instructional advances, which roughly correspond to better content and better method. First, cognitive theoretical models provide detailed representations of expert task performance to use as the targets of instruction. In the details, the goals to which procedural knowledge applies and the strategic processes that are responsible for the organization, coherence, and general execution of the performance are clearly established. With detailed cognitive models as input to instructional systems such as Sherlock 1 and 2, knowledge is directly tied to its uses in the world, and tacit knowledge (including goals, strategies, and assumptions) is made explicit for teaching. Content is thereby richer, more precise, and surrounded by context that establishes conditions of use.

Second, better method has been achieved through a union of modern formulations of skill acquisition and traditional apprenticeship training techniques, such as modeling and coaching (Palincsar & Brown, 1984; Scardamalia & Bereiter, 1985; Schoenfeld, 1985). The common element in both is the notion of skill development as successive stages of increasingly mature performance. Hallmarks of apprenticeship training methods that are consistent with constructivist views include situated learning, external support or scaffolding in the instruction in the form of ideal modeling of the performance, hints, reminders, explanations, or even missing pieces of knowledge to assist the construction of understanding, and carefully sequenced learning activities that are both sensitive to changing student needs at different stages of skill acquisition and robust and diverse enough to foster integration and generalization of knowledge and skill (Collins, Brown, & Newman, 1987). Finally, to synthesize and reinforce the problem solving process, the solution steps are reflected upon, i.e.,inspected, evaluated, and compared to examples of more advanced solutions at the end of each session. The Sherlock tutors were designed with these "better methods" driving the instructional blueprint.

REFERENCES

Bielaczyc, K., Pirolli, P., & Brown, A. L. (1993, April). Strategy training in self-explanation and self-regulation strategies for learning computer programming. Report No. CSM-5. Berkeley, CA: University of California, School of Education.

Bloom, B. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, Jun/Jul 1984.

Chi, M. T. H., & Van Lehn, K. (1991). The contents of physics self-explanations. Journal of the Learning Sciences, 1, 69-106.

Collins, A., Brown, J.S., & Newman, S. (1987). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L.B. Resnick (Ed.). Cognition and instruction: Issues and agendas. Hillsdale, NJ: Erlbaum.

Eggan, G., & Lesgold, A. (in press). Modelling requirements for intelligent training systems. (in press). In S. Dijkstra, H. P. M. Krammer, & J. J. G. van Merrienboer (Eds.), Instructional-design models in computer-based learning environments. Heidelberg: Springer-Verlag.

Glaser, R., Lesgold A., & Gott, S.P. (1986). Implications of cognitive psychology for measuring job performance. Washington, DC: National Academy of Sciences Committee on the Performance of Military Personnel.

Gott, S. P. (1987). Assessing technical expertise in today's work environments. Proceedings of the 1987 ETS Invitational Conference. (pp 89-101). Princeton NJ: Educational Testing Service.

Gott, S. P. (1989). Apprenticeship instruction for real world cognitive tasks. In E. Z. Rothkopf (Ed.), Review of Research in Education. Vol. XV. Washington, DC: AERA.

Gott, S. P., Hall, E. P., Pokorny, R. A., Dibble, E., & Glaser, R. (1992). A naturalistic study of transfer: Adaptive expertise in technical domains. In D. K. Detterman & R. J. Sternberg (Eds.). Transfer on trial: Intelligence, cognition, and instruction. (pp 258-288). Norwood NJ: Ablex Publishing Corp.

Gott, S.P., Pokorny, R. A., Alley, W. E., Kane, R.S., & Dibble, E. (in press) Development and evaluation of an intelligent tutoring system: Sherlock 2 -- An avionics troubleshooting tutor. AL/HR TR-94-XX. Brooks AFB TX.

Hall, E. P., Gott, S. P., & Pokorny, R. A. (in press). A procedural guide to cognitive task analysis. AL/HR TR-94-XX. Brooks AFB TX.

Kane, R. S. (1993). Frankenstation: Functional description and technical data. Unpublished manuscript.

Katz, S. & Lesgold, A. (1991). Modeling the student in Sherlock II. In J. Kay & A. Quilici (Eds.), Proceedings of the IJCAI-91 Workshop W.4: Agent modelling for intelligent interaction (pp. 93-127). Sydney, Australia.

Katz, S., & Lesgold, A. (in press). The role of the tutor in computer-based collaborative learning situations. In S. Lajoie & S. Derry (Eds.), Computers as cognitive tools. Hillsdale, NJ: Lawrence Erlbaum Associates.

Katz, S., Lesgold, A., Eggan, G., & Gordin, M. (1992). Self-adjusting Curriculum Planning in Sherlock II. Lecture Notes in Computer Science: Proceedings of the Fourth International Conference on Computers in Learning (ICCAL '92). Berlin: Springer Verlag.

Katz, S., Lesgold, A., Eggan, G., & Gordin, M. (in press). Modeling the student in Sherlock II. Journal of Artificial Intelligence and Education (Special issue on student modeling), Ed. J. A. Self. Association for the Advancement of Computing in Education (AACE).

Lajoie, S., & Lesgold, A. (1989). Apprenticeship training in the workplace: Computer coached practice environment as a new form of apprenticeship. Machine-Mediated Learning, 3, 7-28.

Lesgold, A. (In press). Assessment of intelligent training systems: Sherlock as an example. In E. Baker and H. O'Neil, Jr. (Eds.), Technology assessment: Estimating the future. (Tentative title). Hillsdale, NJ: Erlbaum.

Lesgold, A., Eggan, G., Katz, S., & Rao, G. (1992). Possibilities for assessment using computer-based apprenticeship environments. W. Regian & V. Shute (Eds.), Cognitive approaches to automated instruction (pp. 49-80). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lesgold, A., Katz, S., Greenberg, L., Hughes, E., & Eggan, G. (in press). Extensions of intelligent tutoring paradigms to support collaborative learning. In S. Dijkstra, H. P. M. Krammer, & J. J. G. van Merrienboer (Eds.), Instructional-design models in computer-based learning environments. Heidelberg: Springer-Verlag.

Lesgold, A., & Katz, S. (1992). Models of cognition and educational technologies: Implications for medical training. In D. A. Evans & V. L. Patel (Eds.), Advanced models of cognition for medical training and practice (pp. 255-264). NATO ASI Series F, vol. 97. Berlin: Springer-Verlag.

Legold, A. M., Lajoie, S. P., Bunzo, M., & Eggan, G. (1992). SHERLOCK: A coached practice environment for an electronics troubleshooting job. In J. Larkin & R. Chabay (Eds.), Computer assisted instruction and intelligent tutoring systems: Shared issues and complementary approaches (pp.201-238). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nichols, P.,Pokorny, R. Jones, G., Gott, S.P., & Alley, W.E. (in press). Evaluation of an avionics troubleshooting tutoring system. AL/HR TR-94-XX. Brooks AFB TX.

Palincsar, A., & Brown, A. (1984). Reciprocal teaching of comprehension fostering and monitoring activities. Cognition and Instruction, 1(2), 117-175.

Pirolli, P., & Bielaczyc, K. (1989). Empirical analyses of self-explanation and transfer from learning to program. Proceedings of the 11th annual conference of the Cognitive Science Society. (Pp. 450-457). Hillsdale NJ: Erlbaum.

Pokorny, R. A. & Gott, S. P. (in press). The evaluation of a real world instructional system: Using technical experts as raters. AL HR TR-94-XX. Brooks AFB TX.

Scardamalia, M., & Bereiter, C. (1985). Fostering the development of self-regulation in children's knowledge processing. In S.F. Chipman, J.W. Segal, & R. Glaser (Eds.), Thinking and learning skills: Research and open questions. Hillsdale, NJ: Erlbaum.

Schoenfeld, A.H. (1985). Mathematical problem solving. New York: Academic Press.

VanLehn, K., Jones, R. M., & Chi, M. T. H. (1992). A model of the self-explanation effect. Journal of the Learning Sciences, 2, 1-59.

APPENDIX I: The F-15 Manual Avionics Test Station

A test station is a large switch, more or less like a telephone exchange. It also contains instruments for measuring electrical energy patterns, such as a digital multimeter and an oscilloscope, and devices for creating patterned energy inputs to the aircraft component being tested. Each test on a box from an aircraft (called a line-replaceable unit (LRU) or sometimes the unit under test (UUT)) involves applying patterned electrical energy to various inputs of the UUT and then connecting various of its outputs to a measurement device. A central section of the test station, called the relay assembly group (RAG) mediates the switching process. The technician sets various switches on the front of the test station to specify a particular test configuration, and then the RAG effects that configuration by energizing relays in giant switching trees. When all the relays are set properly, a signal circuit is created in which electrical inputs go from power supplies and signal generators on the test station, through an active connecting cable array (called the test package, TP), to the UUT, and outputs go from certain pins on the UUT's electrical interface through the switching array to a measurement device. On some test stations, a computer executes a series of tests of the UUT by directly controlling switching relays, but on the F-15 manual station, switching is effected via control settings on the test station's front panel.

When a test station fails, this failure is manifest in some function that the test station does not perform properly. A first requirement in the face of a possible failure is to be sure that the abnormal outcome is not due to a fault in the unit under test, the box from the aircraft. This is the most likely situation - after all the whole purpose of the test station is to reveal faults in aircraft components. Another possibility that must be ruled out is a failure of the test package, the component that connects the UUT to the test station. If both the UUT and the Test Package are operating normally, then the problem is in the test station itself.

The top-level diagnostic strategy would first attempt to isolate the problem into one of two main functional areas. Either the patterned energy inputs are not getting to the UUT, or its outputs are not getting to a measurement device successfully. A single test of the inputs of the test station to the UUT will reveal if the problem is on the signal input side or the output measurement side. The next step is to trace the signal through the pathway, ending with an identification of a component that receives good inputs but has faulty outputs. If this component is involved in the switching process, there are two ways it could be failing. Either it is broken itself, or it is receiving wrong control signals from the switches on the front panel of the test station. In this latter case, the control inputs to the component in question will be wrong, and attention should be turned to diagnosis of the path from control switches to the component now being addressed. On the other hand, if a component has good signal inputs, bad outputs, and good control inputs, then that component is a candidate for replacement.

This global strategy is really a combination of a weak or general method of space splitting, or "divide and conquer", with a specific model of the test station that provides an understanding of the meaningful units of the system that should be the focus of space splitting. The tactics for testing a particular subset of the test station, on the other hand, represent specific knowledge that can be generalized after appropriate experience and perhaps some expert suggestions.