INTELLIGENT TUTORING AND ASSESSMENT
BUILT ON AN UNDERSTANDING OF
A TECHNICAL PROBLEM-SOLVING TASK

Linda S. Steinberg
Drew H. Gitomer

Educational Testing Service
Princeton, New Jersey
November 1996
Intelligent tutoring and assessment built on an understanding of a technical problem-solving task
Linda S. Steinberg and Drew H. Gitomer
Educational Testing Service

Abstract:

In this paper we describe Hydrive, an operational computer-based intelligent tutoring system built to help Air Force technicians develop generalizable skills for troubleshooting hydraulics systems. We use Hydrive as an example of how effective training and assessment is developed from a coherent understanding of a target task and how this understanding can be consistently represented in all aspects of a training/assessment instrument. We show how an organizing principle of expert behavior, active path analysis, is used to inform the design and content of the tutor's system, student and instructional models and the design of the tutor's interface. We conclude with information about Hydrive's field trial evaluation and some thoughts on knowledge acquisition and generalizability.

Introduction:

To equip students to solve real-world problems effectively, training requires a coherent understanding of the real-world tasks students must someday perform once they complete training (Kieras, 1988). The purpose of this paper is to describe Hydrive, an operational, computer-based intelligent tutoring system developed for the Air Force to help maintenance personnel (in this case F-15 hydraulics technicians) develop generalizable troubleshooting skills. We offer Hydrive as an example of how effective training and assessment can be developed from a coherent understanding of a target task and how this task understanding is then used to inform all aspects of the tutor.

Hydrive's development was preceded by extensive Air Force-sponsored research into the applicability of cognitive task analysis to technical skills training and assessment, the identification of technical skills which should be targeted to improve maintenance tasks, and improvement of cognitive task analysis methodology to capture how these targeted skills are deployed in technical tasks requiring cognitive problem solving (Gitomer, 1984; 1988; Glaser et al, 1985; Means and Gott, 1988). The outcome of this research was to adopt a strategy of cognitive task analysis based on the differences between those who perform exceptionally well (experts) and those who perform less well (novices and intermediates). The assumption in this approach to task analysis is that those who perform less well have not, for a variety of reasons, acquired or do not use certain
pieces of knowledge, particularly those which are integrated and applied during actual problem solving (as opposed to tasks which isolate and decontextualize single facets of technical skill). It, therefore, becomes the task of the intelligent tutoring system to create an environment in which both expert and novice behavior can be accommodated, while assessment and instruction target the knowledge and skills that characterize expertise.

We will show how a cognitive task analysis led to a distillation of task understanding into a single overarching, generalizable concept, that of active path analysis. We will discuss the meaning of this concept and how its use allowed us to build a bridge, in the form of Hydrive's student model, between the physical, explicit domain of the aircraft and the inferential, implicit domain of student proficiency. In addition to describing the nature of this bridge, we will discuss how the consistent and coherent application of the active path concept to all aspects of Hydrive's design 1) helped to define a probabilistic student profile of proficiency, 2) determined simulation requirements for the system model, 3) gave structure to the instructional model's curriculum and, 4) provided a realistic and logical troubleshooting model around which the functional flow of the interface could be molded. As an illustration of Hydrive's assessment and instructional design, we will walk through the salient portions of a troubleshooting session. We will conclude with some thoughts on the generalizability of this concept to performance assessments in other domains.

**Background:**

Let us begin with a brief look at the hydraulics technician. The F-15 hydraulics technician maintains the aircraft's hydraulic power system, which generates hydraulic power, and all other F-15 systems which use hydraulic power: the flight controls, the landing gear, the canopy (the transparent bubble through which the pilot enters and exits the aircraft), and the jet fuel starter system for the aircraft engines. The technician is present when the pilot checks out the aircraft just before takeoff and is the first person to be briefed by the pilot immediately after landing. The hydraulics technician is responsible for diagnosing and fixing aircraft problems while the F-15 is still on the flightline, the place from which the aircraft taxis to takeoff and returns after landing. During troubleshooting, the technician has access to technical materials, including step-by-step fault isolation guides, descriptions of general systems and principles of operation, schematic diagrams and step-by-step maintenance procedural guides. The actual repair of defective components is a maintenance shop job and not the responsibility of the flightline technician. The priority in flightline maintenance is to find the faulty component as quickly as possible and service or replace it. The cost of any given flightline maintenance job is measured largely in terms of how long it takes: the more time the technician requires to troubleshoot the problem correctly, the longer the aircraft is not able to fly and the higher the cost.
The hydraulics job, like most flightline jobs, requires social problem-solving. The interdependence of avionic, electrical, hydraulic and mechanical systems means that technicians from different specialties need to work together to find where the failure is located. Thus, the hydraulics technician needs to understand the relation between different aircraft power systems and to obtain and use this information about the performance of other systems to make judgments about the hydraulic systems. The physical demands of job tasks also require that multiple people work together in this position. Not only is equipment heavy, but people may be positioned at different places on the aircraft during troubleshooting.

The team approach can, however, lead to an inappropriate distribution of labor which is detrimental to the acquisition of critical skills. Because a great deal of the hydraulics maintenance job is physically time-consuming (it may take hours to remove a 240-bolt exterior panel to access a component), a division of labor can occur in which some personnel routinely take on the thinking tasks, while other, less-experienced personnel take on the physical tasks. Thus it is possible for a technician to have spent several years on the flightline without ever having developed any real troubleshooting skill.

It was anticipated that intelligent tutoring systems would help produce better job performance by providing a full range of domain problems (from the routine to the very challenging) presented in a simulated context of on-the-job conditions, supported by context sensitive instruction and sequenced to match the current proficiency of the student. Working through an entire problem set not only provides the student with much greater exposure to practical troubleshooting, but it also allows learning in an environment that minimizes unimportant and physically tedious parts of the job (e.g., unscrewing bolts) and emphasizes cognitively challenging aspects. Knowledge acquired in this way can be more easily transferred into actual job situations because of similarities between learning and working environments and because the equipment system itself can be presented in a way that stresses the functional characteristics most salient to effective troubleshooting.

**Cognitive task analysis:**

Research related to human problem solving reveals the necessity of beginning with a task analysis that addresses both the environment in which the problem solving takes place and types and patterns of problem-solving behavior. An analysis of a problem-solving task in a given domain is responsible for laying bear the specifics of issues germane to problem solving in general: 1) essential features of the external environment; 2) internal representation of the problem; 3) the relationship between problem-solving behavior and internal problem representation; 4) how problems are solved; 5) what makes problems hard (Newell and Simon, 1972). In this case, the
cognitive task analysis defined the tutor's domain: troubleshooting F-15 hydraulics problems on the flightline. This task analysis attempted to uncover the nature of the troubleshooting task through the study of how individuals at different levels of expertise gather, represent and apply information necessary to solve a problem. It also attempted to define what makes domain problems hard. An accurate assessment of task complexity is important since the things that make the real task difficult may also be hard to implement in a computational system (Roth and Woods, 1989). The analysis was done in three parts. The first phase consisted of orientation visits to US Air Force base maintenance sites. The second phase was a data collection effort using an analytic methodology known as PARI (Precursor, Action, Result, Interpretation) analysis which was developed in the Basic Job Skills Program (Means and Gott, 1988; Pokorny and Gott, 1990). To begin, a set of 12 problems was generated by expert technicians. These problems were meant to represent a range of type and difficulty typical of a flightline technician's experience. Each of approximately 20 technicians, representing a continuum of expertise from novice to expert, then used the PARI structure to generate solutions to at least two of the problems. In the last phase of the task analysis, experts visited the system developers and provided follow-up advice and feedback.

The primary cognitive task analysis work occurred in the second phase, the collection and analysis of PARI data. Initially, the technician was given a description of faulty aircraft behavior and asked to draw a block diagram of the aircraft system/subsystem in which he\(^1\) thinks the fault occurs. Then the technician was asked to explain aloud the cause of the failure in aircraft function. The term PARI derives from the steps which serve as the problem solving protocol: \textbf{P} stands for the precursor or working hypothesis (e.g., 'I think it may be the hydraulic power source.'), \textbf{A} is the action taken to test that hypothesis ('e.g., I am going to check the gauges to see if hydraulic pressure is abnormal.'), \textbf{R} is the observed result of the action reported to the solver (e.g., 'The gauges show normal pressure.'), and \textbf{I} represents the interpretation drawn from that result (e.g., 'Since the pressure is normal, it's not the hydraulic power source; maybe it's the electrical power source.'.). The interpretation forms the basis for a new working hypothesis (next precursor). The following discussion of the PARI analysis results is based on the annotated example of expert troubleshooting appearing in the Appendix.

The PARI analysis provided two types of information. First, it helped define the nature of expert/novice differences in terms of the most critical skills and knowledge to be addressed in the tutoring system. Second, we were able to see the types of troubleshooting actions and procedures that all technicians used during problem solving.

Expert troubleshooting, typified in this example, included the execution and analysis of multiple dynamic tests of the suspect aircraft system. In this problem, for example, the rudders

\(^{1}\text{F-15 hydraulics technicians are almost exclusively male. Therefore, for purposes of presentation, the masculine pronominalization is used.}\)
were not working. The expert approach involved activating the rudders in different ways, thereby creating more than one path to the rudders. This expert first used the control stick to try to make the rudders move, and then used the rudder pedals to try to make the rudders move. After each attempt, the expert observed the behavior of various components along this path (Step 3). The expert then compared the outcomes of these tests, drew conclusions about which type of component that had been activated might be at fault (mechanical) and where along the path to the rudders the fault was (after the aileron rudder interconnect), and solved the problem (Step 4).

We refer to this form of troubleshooting as **active path analysis** because it involves supplying input to an aircraft system intended to activate some target component(s) of interest (usually the one(s) displaying the faulty behavior included in the problem description) in more than one way, observing the behavior of the component(s) of interest, and then drawing conclusions about the status of components on the path. This fundamental strategic model of troubleshooting was manifested repeatedly throughout the PARI data in expert solutions to problems in all aircraft systems. The use of path analysis within problem spaces has been postulated as a major invariant of problem solving behavior applicable across tasks and individuals (Newell and Simon, 1972).²

Using the PARI data, we analyzed individual differences in analysis of active paths during troubleshooting. We then related the successful analysis of active paths to the types of knowledge that effective troubleshooters bring to a problem.

**Presence of mental models:**

Generally speaking, hydraulic systems consist of a limited number of physically accessible components, each having restricted sets of well defined functions. It is possible, then, for hydraulics technicians to 'know' the systems they have to deal with, where 'knowing' means having some idea, or mental model, of how the components in a given system work together to accomplish a function, like opening or closing the canopy. For the task analysis, the conception of a mental model includes some fundamental ideas related to reasoning about physical systems (Williams, Holan and Stevens, 1983): 1) mental models consist of autonomous objects, each with a set of connections to other objects; 2) associated with each object is a set of parameters and rules specifying its behavior; 3) these models are ' runnable' by means of using the topology and rules to propagate information (modify the model's parameters); 4) mental models can be decomposed.

In order to observe and compare the behavior of components when activated in more than one way, technicians who use active paths in troubleshooting must have explicit 'runnable' mental models of aircraft system operation which they use to direct their troubleshooting actions (see

---
²Even in Newell's and Simon's early work, successful problem solving was characterized by an understanding of the impact problem-solving moves had on constraining the problem space.
Figure 1). These models tend to be accurate representations of how the system works, including the flow of control between components and between power systems and the operation of components within the system. Note that in the diagram shown in Figure 1, the expert has included the very path (including a complete representation of the suspect mechanical portions) he activated in Step 3. (However, because flightline troubleshooting entails diagnosis and replacement, not repair, even expert technicians may not understand the internal workings of replaceable components.) By 'running' these models, or activating paths that reveal component function, experts are able to evaluate the results of troubleshooting actions in terms of this system model and make determinations about the integrity of different parts of the aircraft. As exemplified in Figure 2, novices are able to access, at best, severely impoverished mental models, and, therefore, have no basis for troubleshooting decisions. Individuals falling in between expert and novice frequently evidence incomplete mental models of aircraft systems.

*Attendance to physical clues:*

Active path analysis requires two types of actions: actions that cause the path of interest to become active and actions that inspect the physical behavior of one or more components in this path. Failures of the canopy to open or the rudders to move under emergency conditions, for example, provide immediate and important clues about the cause of the failure. While an expert technician is likely to run a system through various conditions to obtain a better sense of when and how failures occur (creating multiple active paths), a novice may lack even the knowledge of how to test aircraft function under various sets of conditions likely to provide data about the failure.

*Procedural expertise:*

Every aircraft component can be acted on through a variety of procedures. Procedural expertise is characterized by the use of actions appropriate for making observations at various points in an active path. In addition, experts are particularly adept at disabling aircraft systems that may provide backup, or redundancy, for the paths they activate, thereby determining whether large portions of the aircraft system of interest are functional or dysfunctional. Novices are generally limited to removing and replacing components or following procedures specified in a binary decision tree procedural aid called a *Fault Isolation Guide (FI)*.
Functional classification:

In using active paths as a general troubleshooting strategy across problems and aircraft systems, experts are able to identify shared and discrete characteristics of components. This ability to classify components functionally plays an important role in the selection of observation points along an active path together with appropriate inspections. Experts generally have a hierarchically organized understanding of the functional characteristics of classes of components beyond the specific instances occurring on the F-15.

Knowledge of failure characteristics:

Experts use their knowledge of failure characteristics to activate paths that isolate the problem to a particular power system or component type. The fact that aircraft symptoms can be similar or identical in the presence of failure of many different components, or even widely differing types of components, constitutes an area of significant task difficulty. Being able at least to identify failure symptoms with specific types of components gives the expert a head start in formulating a troubleshooting strategy. From there, the expert will bring the appropriate mental model and knowledge of differing functional classes to bear on pinpointing the unique characteristics of the problem at hand.

Access to knowledge:

In their use of active path strategy, experts are able to access, integrate and apply knowledge in a variety of different contexts; that is, across types of problems and across aircraft systems. They are able to construct (from memory or with reference to schematics) the applicable mental model and transfer abstracted, or generalized, strategies and procedures to generate a problem solution.

Flexibility in strategy selection:

Experts exhibit a great deal of variation in their problem solutions. For numerous legitimate reasons, individuals may choose to approach a problem in different ways. While experts may not know all the optimal actions to pursue at a particular point, they possess robust strategies that lead to effective actions. Most strategies involve space splitting of some sort, as in the use of active path
analysis, and require a 'runnable' model of how the particular aircraft system works. Experts usually attempt to isolate the problem to a particular active path first, and then work within sections of that path relating to a particular power system (hydraulic, mechanical or electrical). A frequent exception to this general rule is when an exceptionally cheap action is available that will provide some new information about the system.

Cost/benefit awareness:

Experts try to use strategies that maximize information gained while minimizing cost, where cost is directly proportional to the amount of time required to execute the procedure (dollar value of replacement parts is also a consideration, but subordinate to time). As a rule, they use space splitting strategies, like the creation and analysis of active paths, to rule out large sections of the problem area through the application of relatively inexpensive procedures. The ability to balance cost and information is one of the hallmarks of expertise in this domain. A novice's strategic repertoire is frequently limited to removing and replacing components; cost/benefit judgment does not develop in the absence of strategic options, and the development of strategic options cannot progress in the absence of aircraft knowledge. Technicians of intermediate ability manifest a mixed approach, usually depending on the completeness of their model of how the aircraft works.

The PARI analysis also highlighted the types of individual troubleshooting actions technicians use during problem solving. The choice and sequencing of these actions at various points in the problem characterize a student's strategy. Many of these actions are evident in the PARI example and include:

- powering the aircraft or subsystems of the aircraft
- reading gauges and indicators
- setting switches to create dynamic test environment
- disabling aircraft subsystems and backup functions
- initiating dynamic tests (providing input to controls)
- visually observing the operation of components
- testing electrical function of components
- removing and replacing components

In summary, through the PARI data, the nature of the flightline troubleshooting task was characterized in terms of three interdependent types of knowledge: system knowledge (how it works), procedural knowledge (given a component or a system, what actions can be taken on it),
and strategic knowledge (how to organize system and procedural knowledge into effective strategies, or plans, for finding the fault). Failure to use effective procedures or strategy may not necessarily imply a poor understanding of procedures or strategy. The tendency to engage in certain procedures or strategies is often a function of the structure and completeness of system understanding, rather than the understanding of strategies or procedures in the abstract. Failure to engage in space-splitting strategies like the use of active paths may not, per force, suggest a deficit in strategic knowledge. Failure to use effective strategies may be attributable to one of several factors. First, the troubleshooter may not understand the system sufficiently to be able to 'run' some paths through it. Second, there may not be sufficient system knowledge to know where to observe the behavior of aircraft components along a path (where to split the system). Third, the individual may not have available appropriate actions (procedures) that will effectively collect data about aircraft behavior, especially at the component level. Another possibility is that the troubleshooter is simply unaware of how and when to use a space-splitting strategy like active paths analysis. For those beyond the novice levels, the greatest reason for ineffective problem solving typically is attributable to poor system understanding. For the more novice individuals, there may even be an absence of a general aircraft system understanding that specifies the relationships between power systems. If a technician has exhibited strong strategic understanding on other problems for which a good system understanding exists, then the likelihood is greater that the performance deficit on a new problem is directly attributable to poor system knowledge. Therefore, the task analysis provides a basis for articulating the conceptual interdependencies we assume exist between different forms of understanding.

The results of the cognitive task analysis had major implications for the design and content of the tutor's models and interface.

**Structure of Hydrive:**

Figure 3 is a very simple representation of the standard components of an intelligent tutoring system as organized for HYDRIVE: the system model, representing domain knowledge (i.e., troubleshooting F-15 hydraulic systems); the student model, representing what the student does or does not know; the instructional model providing coaching in the domain as informed by the student model's cost/benefit evaluation of student troubleshooting actions; and the interface, through which the student acts on the system model and receives feedback. In the following sections we describe how the cognitive task analysis influenced and informed development of the respective components of the tutor.
Student model goals.

We begin with student model goals because they embody the fundamental purpose of the tutor, i.e., helping students learn to troubleshoot. The goals for Hydrive's assessment function, realized in the student model, were

- the ability to make claims about an individual with respect to various problem solving abilities within the context of 1) a specific problem; 2) a specific F-15 hydraulics system; 3) all F-15 hydraulics systems; 4) domains other than F-15 hydraulics systems

- the ability to formulate characterizations of student strategy (as evidenced in troubleshooting actions) throughout a problem to serve as the basis for the instructional model's prescription of instruction at a level of cognitive complexity that can lead to increased understanding and successful performance

- the ability to predict the quality of student actions, given a particular problem state and what the student model infers about student understanding

On the flightline, evaluations of a technician's performance based only on the outcome of the problem solving process can be limited to 1) 'That cost too much.' 2) 'That was fast.' 3) 'That's still not working.' This type of assessment provides no basis for improving performance or addressing elements of proficiency because, aside from the outcome, there is no information about the troubleshooting process itself. In fact, a supervisor who is able to take the time to observe a technician's flightline problem solving can offer evaluation and support related to the quality of individual troubleshooting actions leading to a solution (replacement of faulty component), where quality means how much the action tells you about which component might be faulty. On the flightline, and in Hydrive, the quality of actions is considered in a cost/benefit framework, where the most important element of cost is the amount of time a troubleshooting action takes, and the most important element of benefit is the amount of information about the problem the action yields. Assessment based on an examination of the troubleshooting process can yield outcomes throughout, such as: 1) 'What does that tell you?' 2) Can you think of something less expensive than the action you took (i.e., ripping out the engine)? 3) 'You understand how this system works but you don't know how to use that information to plan your troubleshooting.' 4) You need to gain a basic understanding of aircraft function.'

In Hydrive, as on the flightline, troubleshooters express their strategies not through verbal explanations (as in the PARI analysis), but by the sequence and quality of their actions. Actual strategy formation is a hidden phenomenon, accessible only through those observations of a student's actions (Glaser, et al.,1985). These actions are the only manifestations of the higher order thinking required in troubleshooting, in other words, the only evidence of the degree to which the student knows how to use how-it-works knowledge (Kiera, 1988). Given the results of the cognitive task analysis, the claims about a student's proficiency, embodied in the student
profile, would have to be expressed in terms of the novice/expert troubleshooting characteristics we have described. These claims would somehow have to be inferred from types and patterns of actions (strategies) used by novices and experts. The inferences about student understanding made from these actions can serve as both a guide to instruction and a predictive model. This evaluation of actions can begin to build a bridge between the domain of F-15 hydraulics troubleshooting and a profile of the student's troubleshooting proficiency.

To go further, however, we have to address what is meant by the sequence and quality of actions. First, what does quality mean in terms of gaining information about the problem? Second, the quality of any action depends on the context in which it is taken, i.e., the problem itself and the point in the problem solution when an action is taken. The implication for assessment requires that the evaluation of any action include an answer to the question 'What is an optimal troubleshooting strategy at this juncture?'. For example, if you take your car to a garage because the lights don't go on and the first thing the mechanic does is replace the engine, you may think this action inappropriate. If, however, there is smoke pouring out from under the hood and someone observes that radiator is leaking, then replacing the radiator may be required. Replacement is not inherently good or bad. Its relative worthiness is determined by what is known about the problem.

Assessment, therefore, requires three pieces of information: a characterization of the current problem state; a characterization of the student's troubleshooting strategy; and a characterization of the optimal troubleshooting strategy at the same point in the problem. Finally, if, as evidenced in the PARI data, experts use active path analysis as their strategy of choice, how can this concept serve as the organizing principle around which to acquire and analyze information to complete the inference bridge between the F-15 and student understanding? Figure 4 represents the essential problem addressed by Hydrive's student model.

Active path analysis as a framework for the student model:

The diagram shown in Figure 1 was drawn by an expert in response to a requirement to represent the area of the aircraft within which he would be working to solve the rudder deflection problem. The diagram depicts the directional subsystem of the F-15's flight control system. This is the subsystem containing the rudders, which are responsible for moving the aircraft left and right. We think of this as a problem area, or problem space, because this is the area containing all the components that could be relevant to the failure.

When a technician uses the aircraft, or the student uses the system model, to simulate certain aircraft conditions (creates an active path) and then observes the results of the simulation, information about the problem area is presumed available. If, for example, the technician moves the control stick and the rudders move as the technician might normally expect, then an inference
can be drawn that all components involved in rudder operation when activated with the control stick (i.e., in the active path from the control stick to the rudders) are functioning correctly and should be eliminated as sources of the problem. To eliminate them as sources of the problem means to **remove them from the problem area**. In this case, all components not on this active path must remain in the problem area because they are still suspect. If, however, the rudders do not move as normally expected when activated by the rudder pedals, then the technician should be able to make the inferences that 1) some component is not working correctly along the active path from the rudder pedals to the rudders; 2) because the failure lies in this path, any component not on this path can be eliminated from the problem area; 3) components common to both active paths must be working and, consequently, should be removed from the problem area; 4) the failure must be in some component unique to the second active path (this is what remains in the problem area).

As in Step 3 of our PARI data, observation of components at intermediate points along this active path can provide information about subsets of components involved in the operation of the rudders. If an expected output is produced at point x (a mechanical linkage is moving as it should), then an inference can be made that the faulty component does not lie within the subset of components between the point of activation, or control, and the point of observation and that these components can be eliminated from the problem area. By the same token, if an expected output is not produced at point x, then an inference can be made that the faulty component is somewhere between the rudder pedals and the point of observation.

The process of reducing a problem area provides a mechanism through which to gain information about the problem and, therefore, judge the quality of a particular action. Eliminating components tells which ones are no longer candidates as sources of the failure. We can tell what and how many components, if any, were eliminated because of some action taken. The ones remaining in the problem area must then be the focus of troubleshooting. With repeated application of this process, the only thing left in the problem area is the failed component. An action which eliminates multiple components from the problem area is of higher quality than an action which eliminates only one component; this is generally true because the cost effectiveness of the former is greater than the latter in the amount of information derivable from it.

The process of reducing a problem area also defines the type and pattern of actions necessary to effect reductions. For instance, imagine what would happen in our PARI example if the technician continued to activate the rudders in different ways (i.e., cycling repeatedly through the control stick and through the rudder pedals) without making any observations. There could be no reduction in the problem area because, without observations, there is no information available about whether or not the rudders are working. On the other hand, imagine making infinite observations of all components related to the rudders without having activated them in any way. There would still be no way of reducing the problem area because the components haven't been engaged through a technician's actions. Activation and observation are both necessary in order to
effect a reduction of the problem area. To interpret strategy from action, patterns can be matched to the classes of novice and expert behavior defined by the PARI data. Actions required to activate some path in a problem area and make observations within it constitute an expert pattern of problem area reduction. Removing and replacing a component is a novice's way of reducing the problem area. A pattern of behavior for proficiency in between expert and novice consists of a mixture of expert and novice behaviors; that is, at some points in a problem, the behavior may look expert and at other points it may look novice.

The framework for the student model, therefore, had as its target a profile describing the student in terms characteristic of the continuum of expert to novice behavior derived from the cognitive task analysis. A subset of the entire student profile is represented in Figure 5. This profile is updated by aggregate interpretations and evaluations of particular actions. Hydrive interprets the strategy underlying a pattern of action taken by the student, evaluates its quality depending on the problem state, then updates the profile.

Establishing the mechanism for inferring understanding from actions enabled the implementation of Hydrive's models. Fulfilling the goals of the student model within the active path analysis framework necessitated the implementation of three primary components for Hydrive's student model: the action evaluator, the strategy interpreter, and the student profile. Because action evaluation is based on information obtained from simulation of aircraft behavior, we begin with a brief discussion of system modeling in Hydrive.

**The system model.**

For the hydraulics technician, Hydrive's system model appears as an explorable, testable aircraft system in which a failure has occurred. The student uses the system model to simulate various aircraft states and explore the results of these simulations as a means of finding where in the system the problem resides. Task analysis implications for the system model were 1) the model should only be built down to the level of detail required by the task environment: components replaced on the flightline; 2) to accommodate the use of active path analysis, the model would have to be able to simulate individual component behavior as well as overall aircraft behavior under dynamically variable conditions; 3) procedural knowledge would have to include actions (and their relative costs, primarily in terms of time) taken by experts and novices on any given component (individuals in between expert and novice used mixtures of expert and novice actions).

The system model is defined as a set of components connected by means of inputs and outputs. Connections between the components are expressed as pairs of components, the first

---

3 In Hydrive, a low cost action is one that may take a few moments, a medium cost action an hour or so, and a high cost action half a day or more.
being the component producing an output to the second in the pair which receives it as an input. These pairings are called **edges** and are also qualified by the type of power connection (electrical, hydraulic or mechanical). For example, the connection between a rudder and its actuator (the servomechanism which causes it to move) would be *left rudder actuator - left rudder (mechanical)* because the actuator produces a mechanical output which the rudder processes as input. Every component has a small set of possible inputs. For example, the landing gear control handle can be in the *up* or *down* position. The output of a component is controlled by its input(s) and the internal state of the component (whether or not it is failed). Given a set of inputs, the component will produce one or more outputs, the value of which depends on whether or not the component is working. For example, moving the landing gear control handle to the *down* position will mechanically activate a relay, resulting in the creation of an electrical path that energizes the mechanisms associated with landing gear operation, assuming none of these components is failed. A failure may cause no output or incorrect output to be produced. The system model includes a failure capability whereby any type of input to any component can be failed to produce no output or incorrect output. Only one component can be failed for any given problem, a reasonable constraint in this domain. On the other hand, many different failures may produce similar or identical abnormal symptoms in aircraft behavior, creating great flexibility in defining problem complexity and difficulty level.

Every component also has a set of actions (procedures) that can be performed on it. Some components can be set or manipulated (e.g., switches or controls), others can be checked for electrical function (e.g., relays), and others can be inspected visually (e.g., mechanical linkages). All components can be replaced.

The system model processes the actions of the student and propagates sets of inputs and outputs throughout the system (Kaplan, et al., 1993). The student activates the system model by providing input to the appropriate components and then has the option of examining the results of such actions by observing any other component of the system. Thus, a student can move the landing gear handle and then observe the operation of the landing gear. If the landing gear does not move down, the student may decide to observe the operation of other components in order to isolate the failure.

The resulting system model is a representation of the F-15 hydraulics domain that is constrained by the requirements specified by the task analysis: 1) the level to which the model is built is appropriate to the task; 2) all components can be acted upon; 3) the actions observed in expert and novice behavior are available and usable in combinations necessary to recreate characteristic strategic patterns. The use of edges to define the system model allows both a quantitative and qualitative analysis of a problem area. This analysis yields a means of characterizing and reducing the problem area, the mechanism proposed as a foundation of the
active path analytic framework of the student model (that is, problem area reduction as a basis for action evaluation).

*The action evaluator and strategy interpreter:*

In Hydrive, a student's actions are evaluated in terms of the potential information they yield given the current state of the system model. The action evaluator consults the current state of the system model and calculates the effects on the problem area of an action sequence performed by the student on the system model. The action evaluator considers every troubleshooting action from the student's point of view in terms of the information that can be inferred with respect to effects on the problem area.

The action evaluator, in updating its problem area, assumes that the student always makes the correct judgment about whether observations reveal normal or abnormal component states. If, for example, having supplied a set of inputs to create an active path, a student observes the output of a certain component, which the system model 'knows' is normal, then the student is presumed to infer that all edges on the active path, up to and including the component observed, are functioning correctly and, therefore, remove them from the problem area, as the action evaluator does. If the student, in fact, makes the correct judgment about the observation and the appropriate inferences from it concerning the problem area, then the dynamic problem area that the student model and the student hold correspond. If, however, the student decides that the observed component output is unexpected or abnormal, then, at least in the student's mind, all the edges in the active path remain in the problem area, any others are eliminated. At this point, what the student thinks about the state of the problem area (with consequences for what actions the student subsequently deems effective) and what Hydrive knows about the state of the problem area (with consequences for how these actions will be evaluated) begin to differ. While Hydrive deliberately does not impose a single, or 'best', solution for any problem, when, as in the latter case, the student's conception of the state of the problem is erroneous, the difference between the cost effectiveness of subsequent student actions and what the student model deems to be cost effective is likely to grow. Thus, student model evaluations of ensuing student actions are likely to serve as signals to the instructional model that feedback to the student is necessary. In Hydrive, students can use a review function to help compare their own idea of the problem area with that maintained by the student model. The student's task is to reduce the problem area until a diagnosis can be made and the faulty component replaced, or until the problem area contains only the faulty component.

The method for reducing the problem area is generalizable to any system that is comprised of components in which sequential flow of control can be defined. As long as one can make a
judgment about the output state of a component, then inferences can be made about the state of components comprising a subset of the active path, from the point of control, where input is supplied, to the point of observation.

Figure 6 presents a grossly simplified hypothetical problem area for a hydraulics-like system. This system has two points of control (where input is supplied) which both send electrical signals to electrical components A and B respectively. Both of these signals are sent to an electromechanical component which outputs a mechanical signal to the mechanical component. Hydromechanical components A and B operate by receiving the mechanical signal as well as hydraulic power from hydraulic circuits A and B respectively.

In this hypothetical model, a number of active paths can be set up to isolate a fault. By activating point of control A, the entire system other than the path that includes point of control B and electrical B are being tested. If the output from the hydromechanical components is unexpected, then the problem is clearly not associated with point of control B or electrical B edges. If expected output were to be obtained when point of control B is activated, then it is possible to infer that the locus of the fault is point of control A or electrical A, for other than these two component edges, the active paths overlap. Other discriminations can be made by selectively disabling hydraulics A and B and observing changes in the output of the hydromechanical devices.

The strategy interpreter makes rule-based inferences about the student's apparent strategy usage by analyzing the quality of information (i.e., quantity and type of problem area reduction) obtained from the action evaluator. Actual strategy interpretation occurs by evaluating changes to the problem area, which is the entire series of edges belonging to the system/subsystem where the fault occurs. As the action evaluator presents each list of edges that have been removed from this area as a consequence of a troubleshooting action, the strategy interpreter, among other things, looks at how many edges have been removed, what power system(s) they belong to, whether or not the student has created an active path, and whether the component observed was on an active path.

Hydrive employs a relatively small number of strategy interpretation rules (approximately 25) to characterize each troubleshooting action in terms of an inferred student strategy and what the best strategy actually is given the state of the problem area. Because the rules are derived from the PARI data, they include the identification of poor as well as effective strategies.

When a sequence of actions results in new status information about more than one edge in the problem space, Hydrive designates the strategy as a type of space-splitting. Hydrive differentiates between several forms of space-splitting, primary among which is active path splitting, which activates different combinations of components in the observation of a particular system function (as in observing the rudders when operated through the control stick as opposed to the rudder pedals). The other space-splitting strategies comprise variations on the use of active paths. There is power system elimination, which checks out and eliminates as a cause of
failure the power system sources necessary to the creation of active paths (as in checking hydraulic pressure gauges or circuit breakers); and there is **power path splitting**, which is a particular subset of strategies for active paths. The use of this strategy can eliminate series of edges having the same power type (e.g., electrical), locate the failure to a particular power type (as in using electrical backup to replace mechanical function), or isolate the failure to a specific piece of the path (as in Step 3 of the PARI example).

Other troubleshooting actions do not result in space-splitting, but are discrete tests of single components. The most obvious is simply removing and replacing a component and observing whether the change results in a fix to the system. Because an inference can be made only about the output edges of the replaced component and no inferences can be made about other components, a **remove and replace** strategy is expensive both in terms of time and equipment and is recommended only when there is a high degree of certainty that the replaced component is faulty.

A **serial or single elimination** strategy refers to actions that only provide information about one edge at a time. A serial elimination strategy is inferred when one action provides information about one edge and the ensuing action provides information about an adjacent edge. Though the remove and replace is a form of single elimination, Hydrive's identification of this strategy is limited to actions which are observations (visual or electrical).

An **FI** strategy is one in which the student accesses the Fault Isolation Guide three times in a row (three is an arbitrary number used to identify deliberate intention, as opposed to accident) and follows procedures designated therein. While such a strategy is not inherently problematic, it is clear from the PARI data that experts and novices use the FI in different ways. Therefore, the evaluation of a set of actions as an FI strategy will result in probes from the instructional model to ensure that the student understands the effects of actions taken.

Other evaluations do not actually infer strategies, but do make claims about the effectiveness of actions taken. **Redundant** actions are those that do not provide any new information about the problem. It should be noted that some actions are not costly to repeat in terms of time or parts. In fact, experts often will rerun a procedure to replicate and validate a finding. It is only when actions are costly and do not provide any new information that they are considered redundant. **Irrelevant** actions are those in which a student performs actions on components that are not part of any active path through the problem area (system/subsystem of interest) or are not connected in any way to the problem area. Replacing the tires when an automobile won't start is an example of an irrelevant action. An example of a student strategy rule is:

*IF the active path which includes the failure has not been created and the student creates an active path which does not include the failure and the edges removed from the problem area are of one power type, THEN the student strategy is power path splitting.*
The evaluation of the quality of a strategy depends on the problem state at the particular point that the strategy is used. While a remove and replace strategy is judged to be poor quality when the problem area allows for space-splitting, the same strategy will be evaluated as better quality when the potential causes of the problem have been narrowed to a couple of candidate components. Therefore, it is necessary for the strategy interpreter also to be capable of characterizing a 'best' strategy, given the state of the problem.

Hydrive makes use of a strategic goal hierarchy, derived from the expert use of active paths and associated space-splitting strategies, to identify the optimal strategy. Because the goal of Hydrive is to foster troubleshooting strategies that are not only effective within the F-15 hydraulics domain but also generalizable to other systems and maintenance specialties, the strategic goal structure consists of 'weak' strategies, or general, domain-independent methods based on space-splitting. The intent is to allow these 'weak' methods to become stronger through specialization, or practice with specific methods (actions) on specific components. A 'weak' method would turn into a 'strong' method through instantiation, or the application of relevant system knowledge (Glaser et al., 1985). Hydrive's strategic goal structure consists of the following hierarchy:

1) power system elimination
2) active path splitting
3) power path splitting
4) isolate failure within power path
   a) serial elimination
   b) remove and replace

An example of a best strategy rule is:

> IF the problem area contains one or more hydraulic power systems, THEN the best strategy is power system elimination.

A comparison of the student's strategy and the best strategy, as calculated by the strategy interpreter, drives the instructional model which makes the embedded strategic goal hierarchy explicit to the student in the form of prompts, reminders and instructional exercises. GUIDON and NEOMYCIN are two examples of systems with which Hydrive shares the principle of clear articulation of expertise (into abstracted categories of knowledge) as the foundation of student and instructional models. (Wenger, 1987).
The student profile:

Action evaluations are used to update a profile of the student. Hydrive's assessment design takes advantage of advances in probabilistic networks to characterize and assess the quality of a student model through the application of probability theory (Gitomer, Steinberg, Mislevy, 1995; Mislevy, in press; Mislevy, 1995; Mislevy and Gitomer, in press). Essentially it combines the statistical power of probability theory with networks that are structures derived through the cognitive analysis of task domains. The student profile is implemented as a probabilistic network using the ERGO (Noetic Systems, 1993) system. Refer again to Figure 5 for a student profile network that includes only a portion of the flight control network. The nodes and relationships in the network are derived from the PARI analysis.

The PARI data made it clear that troubleshooting proficiency as characterized by the use of active path analysis not only required knowledge of strategies, but also of procedures (manifested in the actions necessary to the creation of active paths and observations along them) and the particular system being explored. The PARI analysis also supported the idea that each of these broad knowledge areas contributing to proficiency could be characterized in terms of constituent parts. The interdependencies evident in the PARI data are represented in the student profile network.

The nodes at the right in Figure 5 are those directly updated through the strategy evaluation. They can be thought of as the points at which troubleshooting actions taken by the student on the system model are brought to bear on the abstracted cognitive characteristics of the task -- in other words, where the observable meets the unobservable. All other nodes can be thought of as constructs which have values determined, in terms of probability distributions for their possible values, by evidence captured by the observables. Once the observables are set by the strategy evaluation process, the remainder of the network is updated based on probabilistic relations among nodes. The network element nodes are updated across actions and problems. There is an increasing level of abstraction and generality of inferences about students as one moves to the left of the figure. Looking at the left side of Figure 5, Proficiency is a parent of System Knowledge, Procedural Knowledge, and Strategic Knowledge.

The probability specification when the network is initially constructed is a response to the question 'given that the student is proficient (strong), what is the probability that the student is strong in each of the respective knowledge areas' and also 'given that the student is not proficient (weak), what is the likelihood that the student is strong in each of the respective knowledge areas?' By moderating the probabilities, one can temper the updating in the system so that multiple pieces of evidence influence any judgments. The relative influence of a parent-child relationship is determined by the relative probabilities. Relationships having strong influence are characterized by child probability values that differ quite a bit for different parent conditions. Less influential
relationships are characterized by child probability values that are more similar across different parent conditions. So, for example, because the PARI analysis showed that expert-novice differences were better described by strategic differences than by procedural differences (even novices have some expertise for different procedures), given a strong overall proficiency, the difference in probability values associated with strong and weak strategic understanding is greater than it is for strong and weak procedural understanding. Increasing estimates of strategic understanding will have a stronger impact on estimates of proficiency than will increased estimates of procedural understanding. Similarly, conditional probabilities of observable actions were initially specified based on results from PARI data (see Mislevy and Gitomer, in press, and Gitomer et al., 1995, for a complete discussion of the setting of the initial probability values).

In a tutoring situation, the student should be learning. Therefore, it is important to know where a student is, not where a student was prior to learning. This required a mechanism to allow for getting the best estimate in the true status of student model variables as the student moves across problems. To this end, a recency strategy was adopted; that is, changes to the student model variables effected by past problems are fractionally reduced at the beginning of each problem so that information from the current problems have more relative impact on belief network estimates than data from past problems (Mislevy and Gitomer, in press).4

In sum, Hydrive's student model uses both local and global understanding of expert/novice differences extracted from the cognitive task analysis to build a hierarchically organized student profile extending across all aircraft systems and strategies. This assessment instrument, therefore, can be used to make claims about student proficiency within a variety of contexts, from the specific problem to overall, generalized estimates. It can also be used diagnostically to direct instruction. Finally, because the student profile is inferred from a theory of performance inherent in the task analysis, it can serve as a verifiable and modifiable predictive model for troubleshooting behavior at local and global levels. The testability of student model adequacy is a feature missing from most intelligent tutoring systems.

Goals of instruction:

While Hydrive's system model functions as a principled discovery world for system and procedural understanding, and its student model makes its evaluations based on an implicit strategic goal structure observed in expert troubleshooting, it is only in the instructional model that all of Hydrive's goals are made explicit. Using Lesgold's theory of intelligent tutoring system

---

4 This recency strategy is not yet implemented.
curriculum design (Lesgold, 1988), Hydrive's instructional model articulates the knowledge
interrelationships revealed in the PARI analysis. Figure 7 shows that the instructional model
contains the same tripod of domain knowledge, but this time in the form of prompts and exercises.

The three constructs (system, strategic and procedural knowledge) represent the highest
level of different instructional viewpoints, each depending on the needs of the student as dictated
by the student profile. For example, one student may have no workable model whatsoever of the
system containing the problem. Another student may have a workable model of the physical
system, but no idea what a troubleshooting strategy is. Viewpoints are carefully decomposed into
goals for each of those constructs and then into lesson objects fulfilling each of the goals. Different
viewpoints can (and optimally should) decompose into the same lesson objects; coherence in
curriculum means that individual lesson objects support all of the viewpoints represented.

The instructional model takes into account three primary findings of the task analysis: 1) impedement system knowledge, or the lack of a 'runnable' system model, is the primary cause of impedement troubleshooting strategies; 2) the presence of a 'runnable' model is manifested by the
use of active path analysis as a major troubleshooting strategy; 3) the presence of this 'runnable'
model is also evidenced by an ability to represent it schematically (i.e., experts, in general, are able
both to produce and process schematic diagrams and rely on this ability during troubleshooting).

Hydrive's instruction focuses on effective system understanding and troubleshooting
strategies rather than on specific procedures to take at a given point in the problem. Ineffective
actions raise doubts about a student's system understanding, which might suggest instruction
targeted towards student construction of appropriate system models. A key instructional strategy is
to help students develop a hierarchical model of system understanding that is a critical feature of
expert knowledge. The claim is that effective troubleshooting strategies are more likely to be
utilized in the presence of such structure.

In order to encourage the acquisition of expert strategies, Hydrive's instructional model is
driven by the comparison of the student strategy and what Hydrive 'thinks' is the best strategy
given the current state of the problem. This comparison determines what, if any, diagnostic
feedback will be delivered to the student during troubleshooting. But because the PARI data
revealed significant diversity in expert troubleshooting, the rules which drive the instructional
model had to be designed to give the student great latitude in pursuing the problem solution; the
instructional model intervenes with prompts or reminders (i.e., diagnostics) only when a student
action constitutes an important violation of the rules associated with the strategic goal structure. As
mentioned before, this is most likely to occur when the student's idea of the problem area and the
student model's representation of same diverge in some significant way. For example, if after
looking at a few gauges and indicators, a student decides to replace a high cost component (which
is not broken), the instructional model responds not only with the reason why this is not a good
idea at this juncture but also with questions the student should consider in deciding on the next
action to take; in this case, it encourages duplication of the failure (creation of an active path) in order to collect more data about the problem.

**Implementation of Instruction:**

Hydrive's instructional model consists of both diagnostic feedback and instructional exercises. Diagnostic feedback is presented to the student immediately upon detection of a serious violation of the strategic goal hierarchy, as mentioned above. The student has no control over this feedback, which consists of reasons why the strategy the student is using is not considered to be the most effective at this point in the problem and suggestions for other, more cost/effective, strategic options. After delivering diagnostic feedback, the instructional model also makes a recommendation for an instructional exercise available to the student. A consideration of the student's proficiency with respect to system knowledge (how the aircraft works) and what exercises the student has already completed are included in formulating recommendations for instructional exercises. However, the actual presentation of any particular exercise is under direct control of the student, who is free to take the instructional model's recommendation, choose a different exercise, or continue troubleshooting without any instruction. This approach is consistent with research results in the area of student control (Corbett and Anderson, 1992; Chabay and Sherwood, 1992) which concludes that 1) students show improved learning when they receive feedback at the point when the error is committed, rather than after some delay, as at the end of a problem; 2) students are frustrated with a lack of control over the learning process. It was felt that intervention beyond diagnosis, or the requirement for instructional exercises at particular junctures, might unacceptably increase the cognitive load in an already complex domain. It is also true that accurate diagnosis of the underlying cause(s) of a particular instance of ineffective troubleshooting is, at best, imprecise. Therefore, in addition to intervening with diagnostic feedback only in response to serious missteps and allowing the acceptance or rejection of an instructional recommendation, we gave students the opportunity to choose any instruction thought to be most needed. To help them, we designed the instructional menu carefully, with categories for system (aircraft) instruction and strategy formation; all exercises have meaningful titles.

Almost all exercises, including those for strategies and procedures, require that students identify, build, analyze and/or predict behavior from schematic diagrams of active paths. This emphasis is intended to foster the acquisition of the skill with which experts use diagrams in applying active path analysis. To foster the acquisition of hierarchical models, the schematics available to the student as part of the technical material used in the troubleshooting task have been organized hierarchically, with increasing detail and complexity as the student moves down through the diagrams for a particular system or subsystem. Text, appropriate for a given level, is available
for every diagram, as is video and text for every component within a diagram. The text describes not only the function of the particular component but also places it within a simple hierarchy of functional classification (e.g., type of valve or switch).

The flow of control within the instructional model starts with the assumption that the student must have adequate system knowledge (a 'runnable' model of the aircraft system) before selecting a troubleshooting strategy. Therefore, a student action that fails to reduce the problem area is first examined in the context of the student profile elements pertaining to system understanding. If these indicate a deficit, instruction is recommended to improve the student's mental model of the physical system. The results of many of these exercises (for example, the 'building' of an aircraft system/subsystem) provide direct evidence of the student's system understanding and cause the related profile elements to be updated. After the point that a student's profile elements indicate proficiency in system understanding, ineffective actions are considered in the context of strategic deficit and instruction adapts to emphasize and encourage Hydrive's strategic goal structure. Success or failure in certain of these exercises continue to update relevant profile elements.

In designing the instructional model's diagnostic feedback and instructional exercises, the question of troubleshooting language had to be resolved. While the verbal protocol of the task analysis revealed patterns of actions of troubleshooters, the troubleshooters themselves did not articulate strategy explicitly; they might say that they were, for example, going to look at particular gauges or indicators, but they did not characterize those actions verbally as a power system elimination strategy. Keeping in mind that Hydrive's mission was to foster development of generalizable troubleshooting skill, it was necessary to elevate strategic concepts and descriptions of action patterns from a level connected to a specific F-15 system (e.g., 'Use the control stick or the rudder pedals to move the rudders') to one that was transportable across aircraft systems and beyond the F-15 itself (e.g., 'Create an active path to the target'). Because support of these generalizable strategic concepts is an explicit goal of the instructional model, we had to create a language for thinking about strategy.

As part of Hydrive's development process, the Air Force provided us with technicians of varying levels of expertise to exercise the tutor. These occasions were in no way formal evaluations; they served primarily as an important development tool in giving us feedback about the system's usability and the technical accuracy of Hydrive's content. Technicians were also asked to address the validity of the strategies defined in Hydrive and the appropriateness of the language used to communicate them. They were asked to verify if, for example, the goal of a particular sequence of actions was to eliminate a particular power system as the source of the problem (e.g., no electrical power source) and if describing those actions as power system elimination was a reasonable way to characterize that strategy. These negotiations were productive and resulted in the language implemented in the instructional model.
These sessions were occasions for various patterns of use, decided upon by the technicians themselves: an expert or a novice might work alone, sometimes less expert technicians worked together, and sometimes an expert worked with less experienced troubleshooters. Any of these permutations proved to be viable. An obvious benefit of collaborative problem-solving was that technicians spoke to each other using Hydrive's strategic terminology; this encouraged the understanding that strategy formation is important and hastened familiarization with the language used by the tutor.

Hydrive's problem set, around which troubleshooting revolves, has been carefully constructed to represent a range of complexity for each aircraft hydraulics system. This does not mean there are both simple and esoteric problem descriptions, but rather that solutions to increasingly difficult problems depend on an increasingly thorough understanding of system function, how components act together to accomplish this function, and how these components can be utilized in an effective troubleshooting strategy for that system. For any given aircraft system, the selection of which problem to work on is under the control of the student. However, the student model requires that specific exit criteria be met before a student is allowed to move from one aircraft system to another: the attainment of an intermediate level of proficiency for that system and the completion of at least ten problems (this number assures that the student will not be able to avoid the difficult problems in the set).

The interface:

Hydrive's interface, through which the student accesses the system model and receives feedback, is designed to provide a reasonably faithful representation of the task environment. The aircraft domain, and the problems themselves, are presented to the student in a realistic way that is logically consistent with the flow of task performance. The environment includes technical materials and auxiliary equipment normally used in flightline troubleshooting.

The cognitive task analysis allowed us to determine the focus of the content of the tutoring system: what goals and actions should be targeted to foster improved task performance. With the focus determined, the second key question is whether the tutor's interface simulates and supports the performance of critical actions in a direct and seamless way: how these critical goals and methods are presented. The how deserves careful attention because it determines how easily the student can manipulate the tutoring system (i.e., perform the tutor tasks) to accomplish the domain-relevant tasks. A tutor with an interface that represents salient task characteristics in a manner not consistent with the real task environment will not be effective in improving performance in that environment (Anderson, Boyle, Corbett and Lewis, 1990).
Hydrive’s interface uses video scenarios of dramatized flightline situations to present its problem set to the student. The student witnesses conversations between the pilot and the hydraulic technician who is with the pilot on the flightline. US Air Force personnel were used for the filming and created the dialogue which explores the symptoms of the failure in characteristic terminology.

Students can orient themselves visually outside the aircraft, and then climb on board any area, as represented by the interface. They can locate and act on any system component and, through video, graphics and audio, see and hear what is happening. They can start one or both engines, shut them down or hook up alternate power sources. They can read gauges and indicators, set switches and initiate aircraft operations. They can call on other technical specialists to perform tests and provide results. Until the problem is solved, all aircraft system behavior presents appropriate manifestations of the fault; after the problem is solved, aircraft operations return to normal. The interface design deliberately omitted the necessity of performing procedures where the task analysis revealed no differences between expert and novice performance (e.g., removal of aircraft panels with a wrench).

At all times during the troubleshooting process, students have on-line access to the complete set of technical materials they customarily use, including fault isolation guides, system descriptions and schematic diagrams. Students can also access Help any time during their session; this provides them with detailed navigational support and gives them basic troubleshooting information as well, such as the types of actions they can perform and types of strategies they can use. A Review function is available for the student to keep track of actions taken and results observed. Review can also be requested to present explanations for Hydrive’s interpretations of all evaluated actions. Background information collected during student sign-on is not used in Hydrive’s decision-making. When students complete a problem they can request a progress review which simply reports the number of problems completed and level of proficiency attained for that aircraft system (entry, intermediate, advanced). It was hoped that this type of progress report would reduce students’ desire to ‘game’ the tutor.

As part of the development process, a GOMS (Goals, Operators, Methods, Selection rules) interface analysis (Card, Moran & Newell, 1983) was conducted to examine the first operational version of the interface (Steinberg and Gitomer, 1993). While the PARI cognitive task analysis was successful in providing information necessary to many important interface design goals (i.e., what the troubleshooting environment should look like and what actions/procedures should be provided), it did not specify completely the relation of user goals to the procedures required by the interface. In analyzing our initial interface design, we found that in order to employ a strategy using active path analysis, the transitions between actions necessary to set up a 'run' of the system model (focused in the cockpit) as opposed to actions necessary to observe the results of a 'run' were both numerous and awkward. This was of particular importance since, as previously mentioned, ease of access to the applicable actions and components could have a profound effect.
on the transformation of 'weak' troubleshooting strategies into 'strong' ones. The resulting changes produced an interface with a high degree of fidelity to the technician's working experience and one which encourages the use of strategically effective procedures based on generalizable, and therefore transferable, strategies like active path analysis. Figure 8 shows how Hydrive's interface supports, and is supported by, the tutor's models.

The consistent application of the same task understanding and organizational principles to the design of Hydrive's interface as well as to its models was the last, and arguably the most important, step in creating a cognitively coherent tutor. Without it, the salient manifestations of effective troubleshooting behavior are lost to the student.

**Conclusions:**

Formal evaluation of Hydrive was conducted by the Air Force in field trials during April-June, 1995. The Air Force was responsible for designing this evaluation, which addressed overall tutor effectiveness and effectiveness/patterns of usage for different Hydrive features. The Air Force is in the process of analyzing and publishing the Hydrive field trial data; but preliminary analysis of these data show that Hydrive enhanced technicians' ability to solve problems requiring strategic troubleshooting knowledge (Hall, Personal communication, August 24, 1995). We are currently engaged in analyzing the field trial data collected for all Hydrive student problem-solving sessions. We are using this data in conjunction with ERGO's inference network to explore the validity of Hydrive's student profile.

The Air Force has not yet come to a final decision about the structure of the learning environment in which Hydrive will be used. One possibility for use being considered is in training center classrooms; another is as continuing education at the bases themselves. In either setting, individual or small group use is viable.

As much as Hydrive is designed to fulfill some very specific goals, namely improving troubleshooting skills of F-15 hydraulics technicians, we believe it has a much broader contribution to make. Most importantly, we present it as an example of a design and development process of an instructional/assessment system that is coherently designed to achieve its goals because of the consistent application of the results of the cognitive task analysis to all elements of the tutor. While the interpretation of this type of data is still more an art than a science, and, therefore, remains an area requiring further exploration, we were able to recognize and define principle patterns of expert behavior which served as a design rationale for the entire tutor. The cornerstone of this behavior was the comparison of aircraft function under variable conditions. Expert troubleshooting consistently included activating portions of the aircraft system in question and making observations of aircraft function. Experts went through multiple iterations of this
activate/observe process (using active paths), each time interpreting results to further circumscribe the area in which the problem could occur.

Hydrive’s interface is structured to facilitate the use of active path analysis by locating and organizing in the cockpit troubleshooting actions which activate different paths in an aircraft system and then locating and organizing elsewhere those actions which observe aircraft function. The interface also seeks to maximize ease of movement back and forth between the cockpit and other parts of the aircraft. A review function is provided to help with the cognitive load flowing from interpretation of many observations under many different conditions. Hydrive’s system model includes all the components and actions necessary to running dynamic tests of aircraft function and then observing that function. Its student model evaluates actions, infers strategy and updates its assessment of student understanding based on an analysis of the effect of student actions on the area in which the problem resides. Student understanding, or proficiency, is defined as a probabilistic network of constructs derived directly from the same patterns of expert troubleshooting and their underlying system, strategic and procedural knowledge structures. Hydrive’s instructional model incorporates this rationale into its feedback by encouraging the use of active paths in the formation of troubleshooting strategy and also provides a curriculum to support both the use of active paths and other space-splitting strategies and the system knowledge required for use of these strategies. We stress that the analysis of the data collected in the task analysis was not just used to define specific portions of the tutor, such as aircraft simulation or expert behavior. The data were used to 1) structure a domain that was both sufficiently detailed and properly constrained; 2) identify, decompose and relate the constructs contributing to successful problem solving in this domain, that is create a detailed and explicit structure represented by the profile of proficiency; 3) given that the assessment task is authentic, find a principled way of defining patterns of student behavior (performance) within the task that could be interpreted and evaluated with well-defined rules in terms of the constructs.

We suggest that basing assessment and instruction on disciplined and explicit task understanding is important in any educational domain, not just limited to those that are computer-based or involve troubleshooting. Regardless of domain or delivery system, there is a critical need to define the conceptual models central to a domain and to understand how successful learners and performers use, evaluate, and adapt those models in solving domain. These models can range from landing gear systems, to models of subcellular processes in biology, to models of supply and demand in economics. Teachers and tutors need to have a clear understanding of the essence of a domain and must help learners develop those essential understandings. We think that following this approach in building bridges between domains and standards (profiles of proficiency) in other subject matter areas, even those not related to physical systems, holds great promise in the evolving area of performance assessment. If a cognitive model of domain task performance in which the interrelationship between essential features can be specified probabilistically, and if
student behavior within the domain can be evaluated with respect to some subset of these features, then this approach could be feasible. Successful generalization of this approach derives from an explicit representation of relationships between task constructs, not from any particular qualities of the features themselves.
References


Acknowledgements

Hydrive has been generously supported by Armstrong Laboratories of the United States Air Force. We are indebted to Sherrie Gott and her staff for their contribution to this effort. The views expressed in this paper are those of the authors and do not imply any official endorsement by any organizations funding this work.
Appendix: Expert PARI data

Initially, the expert is given a fault description and asked to represent the candidate problem space with a block diagram. The figure below is a schematized version of the subject matter expert's (SME) initial representation of the fault description "Prior to taxi, the aircraft had no rudder deflection." This problem was designed so that the cause of the fault was due to the breaking or shearing of a mechanical linkage (the splitter or rudder breakout assembly) that controls the operation of both rudders.

![Diagram of rudder control system]

**Step 1**
P: Check for hydraulic power (all flight controls receive hydraulic power to actuators through circuits on reservoir)
A: Check for Circuit A or B lights in cockpit if pilot has not already done so
R: No lights in cockpit
I: Hydraulic pressure is normal according to indicators in cockpit.

In this first step, the SME wants to ascertain the general status of the hydraulic system. Finding no warning lights in the cockpit indicating faulty hydraulic pressure, the SME concludes that the problem is not due to general hydraulic failure.

**Step 2**
P: Verify problem. Check electrical and hydraulic in dynamic test. Want to make all relevant checks there with pilot present and engines running.
A: Put Anti-Skid on (to check flight controls through stick). Cycle flight controls with stick.
R: All flight control surfaces move except both rudders.
I: Hydraulic power is getting to the flight control system; it is eliminated as a possible cause.
Rudder hydraulic servoactuators are eliminated due to infrequency of both failing simultaneously. The control stick is eliminated as a possible cause since it would have to have no effect on any of the flight controls in order to become suspect.

In step 2, the SME performs a dynamic test of the flight control system to determine the functionality of different components within the system. This test requires that the technician manipulate the system through controls in the cockpit and observe the functioning of flight control surfaces on the exterior of the aircraft. Proper functioning of systems that have components in common with the rudder system leads to the inference that the common components are not faulty. This dynamic test has isolated the problem to a set of components that are associated only with rudder control.

**Step 3**
P: Check pedals to see if they are rigged properly (pedals are controlling input to rudders). Check Aileron Rudder Interconnect for proper mechanical function.
A: Put mule on. (Mule is hydraulic power source for aircraft when engines are shut down). Set Anti-Skid switch to off (to control rudders through pedals). Move rudder pedals. Check mechanical linkage from rudder pedals to Aileron Rudder Interconnect. Check mechanical linkage from Aileron Rudder Interconnect to splitter.
R: All linkages are moving.
I: All linkage from cockpit to splitter (not including splitter) is OK. Aileron Rudder Interconnect is OK since the linkage which goes to it and exits it is working properly.
In Step 3, cockpit controls (the rudder pedals) are being manipulated to determine movement of mechanical linkages within the rudder system. The technician then must check movement of mechanical linkages in the fuselage of the aircraft. This requires the opening of doors on the fuselage to allow observation of the linkages. The SME finds that all linkages up to the Aileron Rudder Interconnect are operational.

Step 4
P: Check splitter in dynamic test for conclusive (more than visual check) evidence of fault in splitter.
A: With another person working pedals, check in Bay 5 for movement of two cables, each leaving Rudder Breakout Assembly and going to a rudder actuator.
R: No movement.
I: Splitter is faulty. Call mechanical technician to replace shear bolt on splitter. Problem solved.

In Step 4, the SME continues to trace the path of mechanical linkages, with cockpit control from the rudder pedals. The SME finds that there is no mechanical movement as output from the Rudder Breakout Assembly and therefore has isolated the fault to this component.
List of Figures

1. Expert PARI diagram
2. Novice PARI diagram
3. Simplified Hydrive model function
4. The work of the student model
5. Extract from student profile network of nodes
6. Hypothetical problem area
7. Hydrive’s curriculum structure
8. Hydrive system diagram
Figure 1. Expert PARI diagram
Figure 2. Novice PARI diagram
Figure 3. Simplified Hydrive model function
F-15 hydraulics troubleshooting domain

How can the use of active paths in navigating the domain be used to infer proficiency?

estimate of technician's proficiency in this domain

Figure 4. The work of the student model
Figure 5. Extract from student profile network of nodes
Figure 6. Hypothetical problem area
Figure 7. Hydrive’s curriculum structure
Figure 8. Hydrive system diagram