Learning Complex Cognitive Skills
With an Interactive Job Aid

CSE Technical Report 694

Eva L. Baker, Principal Investigator
UCLA/CRESST

August 2006

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
GSE&IS Building, Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the Office of Naval Research Award Number N00014-02-1-0179, as administered by the Office of Naval Research. The findings and opinions expressed in this report do not reflect the positions or policies of the Office of Naval Research.

LEARNING COMPLEX COGNITIVE SKILLS WITH AN INTERACTIVE JOB AID[1]

Terry P. J. F. Vendlinski

CRESST, University of California, Los Angeles

Allen Munro, Quentin A. Pizzini

BTL, University of Southern California

William L. Bewley, Gregory K. W. K. Chung, Gale Stuart, and Girlie C. Delacruz

CRESST, University of California, Los Angeles

ABSTRACT

Engineering Duty Officers (EDOs) in the U.S. Navy manage large development and procurement processes. Their initial training is provided in a six week *EDO Basic Course* at Port Hueneme, California. The students, who have higher degrees in one or more engineering disciplines, must learn to make complex decisions that incorporate the uncertainty of future events, and to convincingly present their acquisition recommendations to national leaders. *Expected value theory* provides one framework for making such complex decisions. Students compute an estimated value for each alternative choice by summing the utilities of all the potential consequences of that choice, and weighting those utilities by the estimated likelihood of that outcome.

Using the iRides simulation-training system, we developed a software application that provides a simple interface for examining and presenting the expected values of choices in an EDO context. To support the use of this software tool—the *EDO Decision Aid*—in the context of the class, representations of specific alternative solutions to a class problem were developed and presented to student teams working on that problem. The EDO Decision Aid was designed

---

[1] We thank Dr. Mary Lantgen, CAPT Frank Camelio, CAPT Ronald Crowell, CDR John Hill, CDR (S) Scott Heller, CDR Michael Giaque, LCMDR Robert Phillips, all of the US Navy Engineering Duty Officer School, and the EDO students who participated in this study. We also wish to thank Dr. Zenaida Aguirre-Muñoz who assisted with early parts of this study and Ms. Joanne Michuye, who assisted in the work.

to record student actions, including the selection of alternatives, the setting of utility values and estimated probabilities, and the setting of decision thresholds.

Analysis of student actions allows us to compare both the time students spent using the tool to arrive at their final recommendation and the range of alternative solutions they considered. Lag Sequential Analysis then allows us to analyze the order of events in the student's solution process. Two groups can be distinguished: those who approach the acquisition process from a global perspective; and those who considered utility and uncertainty to make a case-by-case determination of value. How instructors might use this information to make instructional changes is explored.

Introduction

Two threads of research on teaching and learning are particularly relevant to the work presented here. Both (1) *the context—tools and environment—in which learning occurs* and (2) *the self-construction of knowledge by students* have emerged as key to successful learning and instruction.

The Importance of the Learning Context

Over the last 40 years, research in cognitive science (e.g. Richards, Jepson, & Feldman, 1996; Cole, 1996) has increasingly shown the importance that the "culture" and the organization of the environment play in developing a learner's "expert" thinking and ability to use available information in problem solving. De Groot (1965) found that the ability of chess experts to recall *realistic* chess game patterns far exceeded the ability of novices; however, there was little distinction in novice / expert ability to recall nonsensical patterns.

Sternberg's (1999) research suggests that strategies are differentially effective depending on the context of the problem (p. 53). Apparently, the context (including the cultural tools at our disposal) often affords useful cognitive aids that encourage or support expert-like behavior. Loewen, Shaw and Craik (1990) found that older adults effectively exploit artifacts in their environment to compensate for memory loss by using external memory aids, and Norman suggests this is not only an important trait for the old, but is something that makes us characteristically human (Norman, 1993).

Glaser (2000) argues that structured knowledge is not just a consequence of the amount of information received, but reflects exposure to an environment for learning where there are opportunities for applying that knowledge through problem-solving.

This not only suggests that instruction must impart knowledge, but that it must also provide opportunities for the real-life application of that knowledge if the structure associated with domain expertise is to develop. In fact, research suggests that students who learn in a problem-solving context are far more likely to spontaneously use (Bransford & Schwartz, 1999) and transfer (Mayer & Wittrock, 1996) what they have learned than are students who are expected to learn solely from the presentation of facts These observations suggest that the integration of practical problem solving into the instructional setting would have positive effects on learning. Shepard (2000) suggests they would also inform our assessments of "expertise" as well. Given that such tasks better represent actual problem solving in a discipline and are aligned to the context in which knowledge being studied will ultimately be applied, authentic problem-solving tasks should improve the accuracy of our evaluations and inferences about the expertise of novice learners (Pellegrino, Chudowsky, & Glaser, 2001; Wiggins, 1993).

This is not to say that the introduction of *any* problem solving experience or *any* basic tool into a course of instruction will be intrinsically beneficial. First, there exist practical tradeoffs when employing problem solving in instruction, and second it is not always desirable for learners to face an unbounded area of study. Quellmalz and others summarize what have been major problems with performance-based assessment activity. They find such tasks to be time-consuming for an already time constrained curriculum, costly to develop, and their dissemination to classrooms to be logistically demanding (Barton, 1999; Edelstein, Reid, Usatine, & Wilkes, 2000; Quellmalz, Schank, Hinojosa, & Padilla, 1999; Shavelson, Gao, & Baxter, 1993).

Second, it is "essential that the [instructor] provide structure when students engage in complex problem-solving using computers" (Sivin-Kachala & Bialo, 1998). Problem solving, by its nature, requires content knowledge as well as experience. Left alone, novice learners can require long periods of time to gain the content knowledge necessary to make good problem solvers, and the frustration this can produce may make students less, rather than more, likely to develop expertise in a knowledge domain. Consequently, Airasian and Walsh (1997) argue the need to strike a balance between instructional formats such as lecture and student problem solving, and Lesgold (1987) suggests the need to use simple inputs and more familiar contexts in problems in order to minimize the load on a novice's working memory.

The objective, then, is to construct realistic problems in contexts that are both familiar and meaningful to students, but that are not so cognitively complex as to overload a student's working memory. To accomplish this objective we must consider what aspects of student learning might benefit from practical problem solving, what aids or tools may facilitate that learning, and how complex to make such a problem. Matching these considerations with student skill development, however, is difficult for teachers (Sutherland, 2002), Tools must be flexible enough to accommodate various types of similar problems and student ability levels, and allow instructors to adjust problem complexity when necessary (Baker, 2002).

Self-Construction of Knowledge

In addition to the contextual aspects of the learning environment, the degree to which the student constructs his or her own understanding of a domain has proven increasingly important to learning and the development of expertise. Although Piaget (1964) felt "learning is provoked by situations" (p. 8), more recently others have argued convincingly that learning is more spontaneous in the sense that the teaching actions do not "cause" the learning. Students self-regulate interactions, so it must be the goal of the instructor to bring forth students' spontaneous schemes (Steffe & Thompson, 2000). Kilonsky (2002), among others, argues that this sort of "active student" achieves better retention and exhibits flexibility, inquiry skills, and higher-order thinking.

There are important caveats. Merely learning *about* a skill is not enough. Brown and Duguid (2000) cite Gilbert Ryle as making the distinction between "know[ing] that" and "know[ing] how". For Ryle, learning about something involves the accumulation of "know that" – principally data, facts, or information. Learning about does not, however, produce the ability to put "know that" into use. This, the authors argue, calls for "know how". And "know how" does not come through the mere accumulation of information. If it did, "know that" and "know how" would, in the end, be indistinguishable. In other words, if one could build up enough "know that," they would become an expert practitioner. We learn how through practice. And, similarly, through practice, we learn "how."

This does not mean that learning and practice should be entirely unstructured. In fact, in their early studies of cognition and learning, Mann and Jepson (1993), and others (e.g., Richards, Feldman, & Jepson, 1992; Richards et al., 1996; Watanabe, 1985; Witkin & Tenenbaum, 1983) suggest it is the orderly structure of the world that allows humans to learn in the first place. But, "structured knowledge is not just a consequence of the amount of information received," observe Glaser and Baxter (2000), "but reflects exposure to an environment for learning where there are opportunities for problem solving, analogy making, extended inferencing, interpretation, and working in unfamiliar environments requiring transfer.” Self-explanation is an important consideration when discussing ways to make learning effective. Left to their own devices, however, students often develop superficial explanations or otherwise rationalize their lack of deep understanding (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Scardamalia & Bereiter, 1993). In fact, the research of Ge and Land (2002), as well as that of Lowyck and Poysa (2001) and others (for example, Chi & Bassok, 1989; Renkl, 2002), argues that students must be forced to make their ideas explicit, justify their conceptual representations, reflect on different ideas, and identify the underlying domain principles in order to go beyond the mere "illusion" of understanding. While some have used stem questions (King, 1994) or question prompts (Camacho & Good, 1989) to deepen the analytical skills of students, we have taken a different approach: integrating hands-on problem-solving and student discussion in a learning environment.

Given the importance of both the learning context and the active student participation for deep understanding in a conceptual domain, we worked with the US Navy to develop a computer-based tool that would integrate well with an authentic operational context, leverage instruction in an existing school curriculum, provide the structure to facilitate hands-on learning, and promote in-depth discussion during the completion of an authentic problem-solving task. Ultimately, our goal is to improve student understanding of a domain, but as an initial “proof of concept,” we also wanted to isolate the ways students thought about their learning in the chosen domain.

Engineering Duty Officer School

The Office of Naval Research (ONR), through its Capable Manpower Future Naval Capability (FNC) program, funded the National Center for Research on Evaluation, Standards, and Student Testing (CRESST/UCLA) and Behavioral Technology Laboratories (BTL/USC) to develop models and tools for use in Navy and Marine operational environments. One of the environments selected by ONR was the Engineering Duty Officer (EDO) School. Given that the developing literature suggests the importance of both the contextual tools and explicit explanation in developing deep understanding, we created automated tools to help evaluate the process EDO students used to develop their frigate acquisition plan and resolve a particular acquisition problem. The Navy’s Engineering Duty Officers (EDOs) manage large-scale development and procurement processes. During their initial training, EDO candidates are taught about making complex decisions as part of project risk management. The students, who have advanced degrees in one or more engineering disciplines, must learn to make complex decisions that incorporate the uncertainty of future events, and to convincingly present their acquisition recommendations to senior Navy officers for approval. During the Basic Course, students are given a variety of techniques for mitigating project risk and for making complex decisions. Exercises are conducted in which three teams of approximately six students each analyze risks in assigned projects and make formal presentations to boards of reviewing officers, who are, in this case, faculty members of the EDO School.

When we approached the EDO School faculty about their needs, the faculty members asked for help in training and assessing decision-making skills in the context of the assigned project exercises. In particular, during the final exercise, students are asked to address a mid-procurement project crisis—the vendor of an important ship system (the Refueling at Sea system, or RAS) has decided to discontinue providing that system. Student teams must determine and evaluate possible solutions and present their recommendations to the review board. Our task was to decide what could be done to make this experience one that enhanced the acquisition of the decision-making skills taught in the course, and to find ways to assess the students' application of those skills.

In addition to assessing student knowledge of risk-management and the federal military acquisition process, several related topics presented to the students in the course drew themselves to our attention. One of these was the topic of multi-attribute measures of utility. Early in the course, students are presented with an example of choosing a restaurant for dinner. Four possible restaurants are considered, and each is given a simple numerical score on such attributes as nearness, expense, atmosphere, and food quality. The concept of weighting utility scores differentially is also introduced, and a simple Excel worksheet for computing the "best" restaurant outcome based on weighted attribute scores is presented. At this stage of the course, the students have been exposed to these concepts:

1. Multiple components of utility (attributes)
2. Weighted attribute values
3. Use of computer-based tools to support decision processes

Later during the course, the faculty briefly introduces Expected Value Theory as a more sophisticated framework for making such complex decisions. In addition to estimating an outcome value for each alternative choice (by summing the weighted utility values of all the potential consequences of that choice), the students also make estimates of the probability of each outcome, given the preceding choice. The expected value of a decision is computed by summing the probability-weighted estimated outcomes of each decision. Although it would be possible to repeatedly make such estimates of probability and utility values and to repeatedly compute expected values by hand, this task clearly would benefit from the use of computer-based support tools. At this stage of the course, the students have also been exposed to these additional concepts:

1. Alternative decision outcomes can be assigned estimated probabilities of occurrence.
2. Expected values can be computed from estimated probabilities and the sums of weighted outcome utilities.
3. Decisions can be made based on expected value analyses.

We have built an experimental tool based on the six concepts just described. The tool has two purposes: to contribute to the students' understandings of these topics, and to provide a natural, problem-centered task for collecting data for assessment. The tool was to be simple, so that students could learn to use it very quickly, and it was to be flexible so it could fit a variety of course contexts (from choosing a restaurant to acquiring a frigate). This would apply the tool to the RAS problem during the last project exercise of the course. In addition, it was decided that the tool would be delivered to the school populated with appropriate content to facilitate its intended uses in the course. This content would include a simple version of the tool applied to the restaurant decision example, and would make it possible to use a simplified form of the tool when multi-attribute utility concepts were introduced early in the course. A simple example— selecting a digital camera—was developed for use in introducing the concept of expected value. The RAS system decision was implemented in the tool, so that students can focus on the estimates they have to make, rather than on the mechanics of authoring every aspect of the alternatives from scratch using the tool.

Method

CRESST/UCLA and BTL/USC designed and developed a computer-based tool, the *Decision Aid Tool* or DAT) to help develop and assess the decision-making skills of students at the EDO school. In addition, to better understand the impact of this tool on its users, we asked each student to complete a short student survey at the end of the course.

Decision Aid Tool

The DAT was designed as a joint project of CRESST and BTL. The technologies used to implement the Decision Aid Tool (DAT) were VIVIDS (Munro & Pizzini, 1998; Munro, Surmon, Johnson, Pizzini, & Walker, 1999; Munro, 2003) and *iRides Author* (Munro, Surmon, & Pizzini, in press). This tool was designed to let training developers create interactive graphical simulations and training in the context of those simulations. The iRides program can deliver the training specifications as a Java application, or over the Web as an applet or a Web Start application. The behavior specification language of iRides is sufficiently expressive and powerful that it was possible to implement a real software tool for aiding decision making using expected value theory.

The tool was developed in three phases, which resulted in three releases of the DAT: prototype, version 1, and version 2. After each of the first two phases, student usage and instructor comments led to significant revisions that appeared in the subsequent release. Some of these modifications were designed to make elements of the user interface easier to learn and to use, to correct algorithmic errors, and to improve data reporting. In addition, however, a number of changes were made to the tool to bring it into compliance with the specific content and within the context of the EDO Basic Course. Examples of this included restricting attribute utility values to integers between 1 and 5, and including three default attributes of outcome utility: cost, performance, and schedule.

Using the DAT

In this discussion, the behavior of version 2.0 of the DAT is described. The data collection employed version 1.0, but the major differences in 2.0 are not relevant to the core issues of operation sequencing in the usage of the tool. The primary difference in version 2.0 is that users are not limited in the depth and breadth of the decision trees that can be authored. In addition, the graphical user interface of 2.0 is improved by the use of Java Swing interface objects in place of certain authored iRides GUI objects.

If an author begins to develop a decision analysis from scratch, the initial display shows only a root decision node and one simple choice branch, as in Figure 1. A pop-up menu is used to select among the commands that display when a node is clicked. On the root node, the options are "Edit Label" and "Create Choice." Create Choice is used to add a new element under the root node, a new decision choice. Other nodes have a "Delete" option, but the root node cannot be deleted.

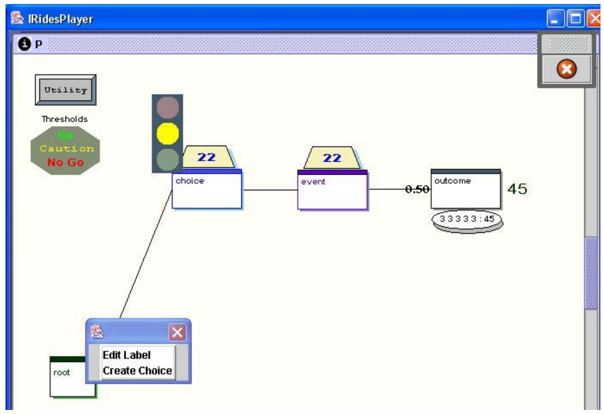

Figure 1. 'Empty' DAT model

Authors re-label the nodes to reflect the choices in the context being analyzed and the possible outcomes of decisions. They can also create new nodes, including additional choices, events, and outcomes. At the development point shown in Figure 2 the original nodes have been relabeled and the author has created two possible outcomes for the first evaluation: a good result and a poor one. For a given decision domain, outcomes have appropriate *attributes*. Authors can enter the names of the attributes that apply to the decision that is being analyzed. Clicking on the *Utility* button opens the Attributes Definition interface (Figure 2) In the original, empty DAT document, there are five utility attributes named "a" through "e."

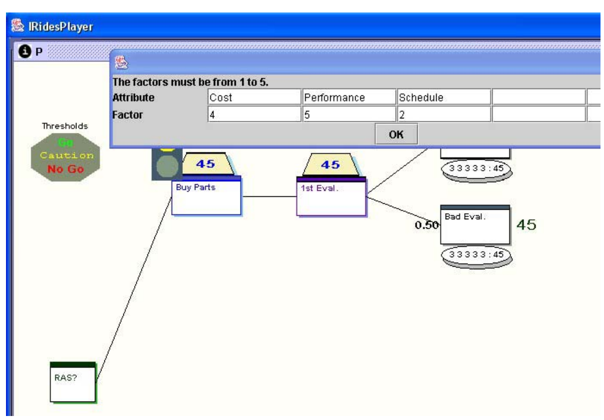

Figure 2. Renaming nodes, defining attributes

Initially, each defaults to an intermediate *factor* of 3. These factors are the weights by which actual attribute values of particular outcomes are multiplied to compute the total value or utility of each outcome.

When the Attributes Definition interface is closed, the attribute values below all outcome nodes are updated, if any attributes have been deleted or if new ones have been added. Because the original attribute names "d" and "e" were deleted, in Figure 3 there are only three attribute value numbers below each outcome, although there were five in the earlier figures.

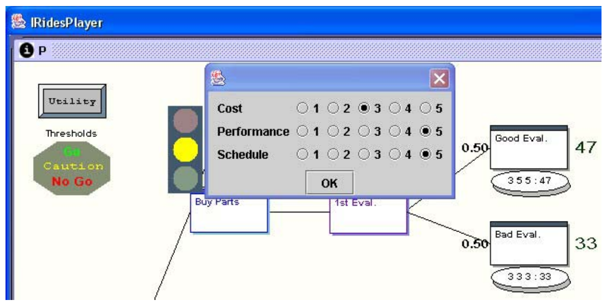

Figure 3. Setting attribute values

Clicking on the attribute values of an outcome node opens the Attribute Settings dialog. For each outcome, the user can specify how good or bad the result will be in terms of each attribute. In the case shown here, the Cost result will be neutral (3) if the parts inventory is purchased and a good evaluation results. The Performance will be excellent (5), and the Schedule will also be excellent, because roughly half of the planned production run will be completed. As these values are selected in the dialog, the numbers change in the outcome's ellipse in the main screen, and expected values are also automatically recomputed.

Not every outcome of a post-choice event is equally probable. The expected value of a choice is dependent not only on the utility of resulting incomes, but also on the probability that those outcomes will occur. Estimated probabilities are shown as numbers to the left of outcome nodes. When new outcomes are first created, they are equally likely. (Note the .50 values to the left of the outcome nodes in Figures 1-3.)

Clicking on a probability opens a Probability slider. In Figure 4, the author has decided that there is a 4% chance of a poor first evaluation after making the "Buy Parts" decision. As the slider is dragged to a new value, the corresponding alternative outcome's probability is automatically altered so that the numbers sum to one. (If an event has three or more possible outcomes, the probability of each outcome must be set manually.)

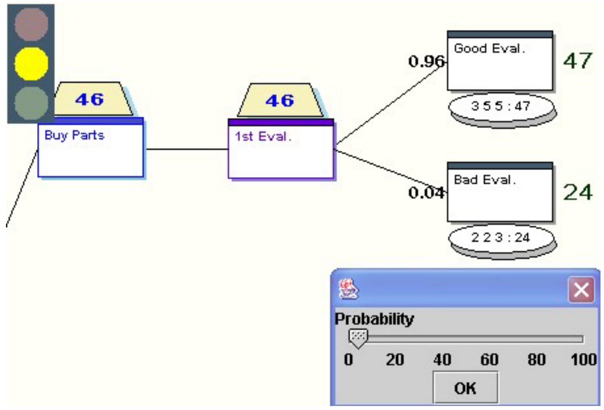

Figure 4. Setting outcome probabilities

By continuing to add choices and outcomes, editing the utility values, and specifying estimated probabilities, a user can develop a rich representation of many aspects of a problem. In Figure 5, the values selected by the user do not result in large differences in the expected values of the choices analyzed. The traffic light signal shown to the left of each choice node reflects the "go–caution–no-go" presentation approach advocated in the EDO Basic Course. In each case here, the three choices are marked here with the yellow "Caution" symbol at the center of each stoplight.

It is possible to manipulate the thresholds of the signals using a pair of sliders. Clicking on the button labeled "Go–Caution–No-Go" reveals the slider interface, as shown at the left in Figure 5. Depending on how the students set the thresholds, all, some, or none of the possible courses of action they propose may produce "acceptable" outcomes.

The RAS Partial Analysis

When students begin working on the Refueling At Sea problem, they open a DAT model that includes three obvious choices (the three shown in Figure 5) but with all outcomes having equal probability and equal utility. As previously discussed, this provides some basic structure to the RAS problem. Students may modify these initial parameters to create more nearly complete analyses. They can also delete choices that they believe are not worthy of consideration, and they can devise and insert new procurement options of their own design.

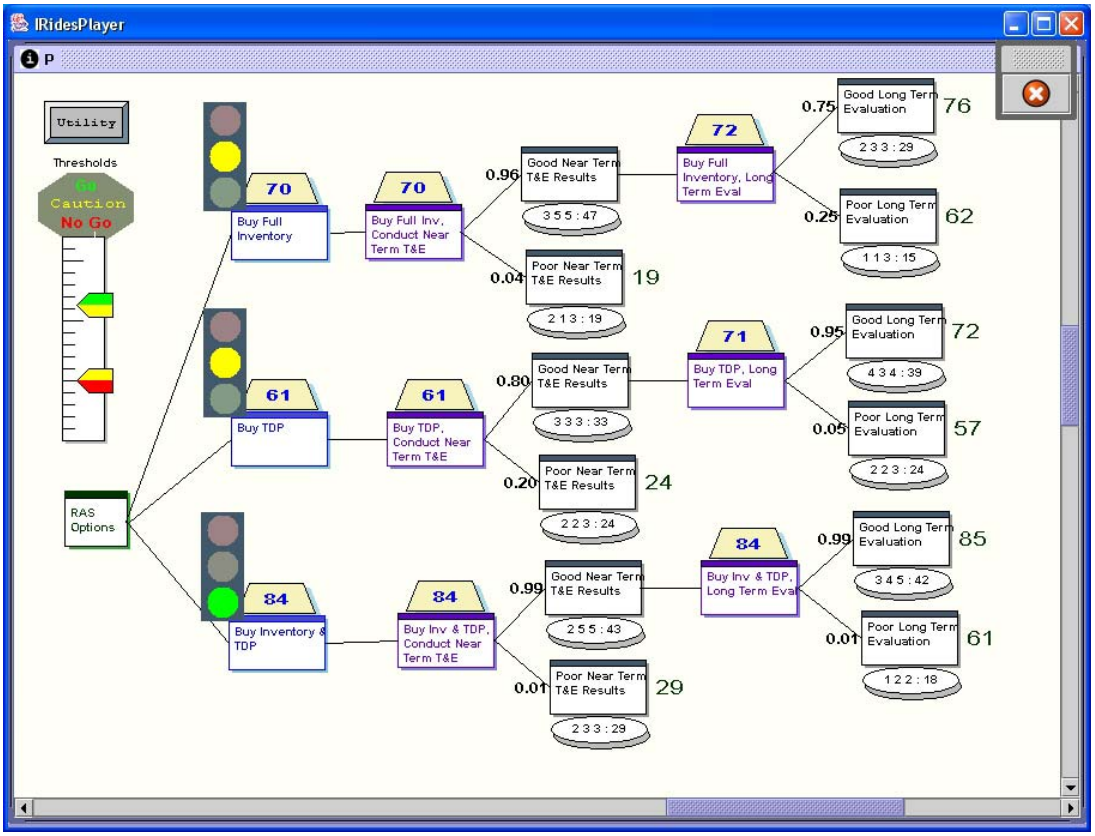

Figure 5. A complete *Refueling at Sea* analysis using the tool

Recording DAT Usage Data

There are seven types of events generated as students use the tool (create a new course of action, delete a course of action, label a course of action, weight the overall importance of utility attributes relative to one another, set the probability that each outcome will occur, set the utility attributes of each outcome, and set the threshold values). The system also generates a "stoplight" event whenever a student action causes the expected value of a possible outcome to cross the student-determined acceptability threshold. When the Decision Aid Tool is used, it records each event in an electronic "clickstream" file. Each file entry includes a student identifier, the date and time of the event, the action that generated an event, the target of the event, and the value assigned to the target of the event. For example: "EDO 33, Monday Feb 09 2004 14:45, label option 1, 'Buy Full Inventory, $20 M.'"

Because students can generate an unlimited number of procurement actions and can evaluate an unlimited number of future decisions made about each action, we cannot specify in advance objects students will create or how they will manipulate those objects before each use of the tool. Consequently, it is difficult to evaluate a student solution by comparing it to a single correct solution or to compare one student solution to another student solution. Even if we defined the largest solution space developed by any student to date as the basis of comparison, the number of degrees of freedom combined with the small number of clicks on less often chosen targets would make meaningful analysis difficult. Finally, the instructors at the EDO school have indicated that they have no present need for such a detailed analysis.

The number of event types generated by a student is fixed at 7 regardless of the number of objects a student creates. This allows us to reclassify granular data in terms of categories. Classifying events at this macroscopic level has the advantage of reducing the degrees of freedom to a level that permits data analysis.

Analysis of Decision Aid Tool Data

After reclassification, we performed a frequency analysis of the clickstream data. This analysis allows us to determine how much time a student spent using the tool, the total number of interactions (clicks) a student had with the tool, and how those interactions were distributed among the eight event types. This analysis is particularly important for identifying students who concentrate on (or ignore) specific aspects of a decision. For example, students who do not adjust the probability values of an outcome may not be considering the impact of the uncertainty of forecasts in their decisions. Frequent changes in utility or probability values may mean students are unclear about these concepts, are performing a sensitivity analysis, or are trying to justify a specific outcome. Frequency analysis also allows us a comparison of how various students navigated the problem space and how their use of the system relates to a model of interaction. Finally, frequency analysis provides us a guide as to which event types occur so infrequently they can reasonably be ignored in subsequent sequential analysis. For example, we suspected that creating new decision paths, while an important characteristic of some solution strategies, would occur so infrequently compared to other click events that their presence (or absence) alone, rather than how often they occurred, would make a solution strategy distinctive.

Frequency analysis, however, has limitations. It is only able to tell us that a decision maker manipulated certain parameters in the problem space. It tells us nothing about the order in which those events occurred. Consequently, frequency analysis allows us to say little about the process that an EDO student took to reach his or her final decision. It is conceivable that the frequency of clicks could be identical for two different student groups, but that the patterns of use are entirely different. We therefore applied sequential analysis to the EDO data. Although we considered only pairs of clicks, the methodology could be extended to larger action sets in order to determine the significance of click triplets, sets of four clicks, and so on.

Pearson's Chi-square ^{2} was calculated for the results of each student group to determine if prior and antecedent clicks are significantly associated (or are independent of one another) at $\chi = .05$. Because of the limitations of the Chi-square statistic when working with a small sample size, we further limit our investigation to those events that occur more than five times in the data, again as suggested by Bakeman and Gottman (p.145). We then computed z-scores (adjusted residuals) for each two-event sequence to identify those transitional probabilities that are significant at the .05 level. Given the sensitivity of z-scores and Chi-square results to sample size, we compute a Yule's Q statistic to compare the magnitude of an effect for each click-pair across group samples. For an extended discussion of methods, see http://btl.usc.edu/edo/toolResults.doc.

Student Survey

In order to validate our inferences about the student beliefs that prompted certain behaviors and to assess the impact of the tool on learning, we administered a 15-item survey to each student. The survey asked them to estimate their knowledge about the domain and the impact of the DAT. Results were correlated with the results obtained from the analyses described above.

Results

The final evaluation of students in the EDO course requires each group of four to six students to present their proposal for the U.S. Navy to acquire a new frigate. The presentation is structured as a Milestone B review and requires each group to consider the cost, performance, and schedule of each major subsystem in the program. One of the subsystems that the students must address in their presentation is the RAS system, how they plan to respond to programmatic changes and uncertainty, and how they will generate and evaluate alternatives in order to accommodate these changes. The RAS problem is given to them only 48 hours before their final presentation. Because other course requirements take up a large portion of the students' schedules, they have limited time to develop new presentations. Consequently, most groups develop their proposals in sub-teams. As a result, only one or two team members generally used the

Decision Aid Tool when evaluating the group's solution to the RAS problem. We collected the data for this analysis from two classes of EDO students. Groups 1–3 used the tool in 2003, and Groups 4–6 were students in a 2004 class.

Decision Aid Tool Frequency Data—Time and Rapidity of Interaction

The data in Table 1 describe the frequency of each type of event generated by students using the Decision Aid Tool, as well as the length of time students used the tool in preparing their presentation. On average, the groups generated events in the tool every 2 or 3 minutes; however, the total number of interactions with the tool and the total amount of time each group used the tool varied quite dramatically. Groups that spent less time with the tool interacted with the system more frequently than groups that spent more time (3.3 clicks per minute on average as compared with 2.5 clicks per minute) and the difference was significantly different between the three groups that used the tool in 2003 and the three groups that used the tool in 2004 (t = -3.3, p = .03). The frequency of each click type also varied between the student groups.

Table 1.

Frequency Distribution of Events by Group

	Group 1		Group 2		Group 3		Group 4		Group 5		Group 6	
Event type	Freq.	% of total	Freq.	% of total	Freq.	% of total	Freq.	% of total	Freq.	% of total	Freq.	% of total
Utility	157	57%	84	51%	33	41%	89	50%	52	66%	81	48%
Stoplight	52	19%	39	23%	19	23%	30	17%	6	8%	54	32%
Probability	34	12%	13	8%	9	11%	18	10%	12	15%	16	9%
Thresholds	18	7%	11	7%	12	15%	14	8%	0	0%	8	5%
Labeling	2	1%	8	5%	4	5%	16	9%	5	6%	3	2%
Weights	3	1%	7	4%	0	0%	0	0%	3	4%	3	2%
Total clicks§	275	100%	166	100%	81	100%	177	100%	79	100%	170	100%
Total time	96 mins.		75 mins.		32 mins.		55 mins.ξ		25 mins.		47 mins.ξ	
Clicks/min.	2.9		2.2		2.5		3.2		3.2		3.6	

ξBoth Group 4 and Group 6 had periods of inactivity (20 minutes and 87 minutes, respectively) in their interaction with the DAT. These periods have not been included in the total time value for either group.

§Total clicks includes create and delete event clicks, but these are not disaggregated in the above table.

Decision Analysis Tool Frequency Data—Utility, Stoplight, and Probability

In all EDO classes, three or four of the eight possible events account for the vast majority of activity in the decision-making process. Given that the stoplights indicate a "go–caution–no-go" decision for each branch of the decision tree, and that utility, probability, and thresholds are the three main determinants of the stoplight colors, this is not surprising. However, since changes to utility, probability, and thresholds each impact a determination of which decision path is optimal, it is surprising all but two groups of students focused on utility significantly more often than would be expected if their actions were distributed in proportion to the click areas in the problem space (Group 1: $\chi^2 = 14.45$, $p < .001$; Group 2: $\chi^2 = 19.2$, $p < .001$; Group 3: $\chi^2 = 52.15$, $p < .001$; Group 4: $\chi^2 = 20.4$, $p < .001$; Group 5: $\chi^2 = 4.38$, $p = .11$; Group 6: $\chi^2 = 7.57$, $p = .02$). For an expanded discussion see http://btl.usc.edu/edo/toolResults.doc.

Groups 3 and 5 were atypical in setting threshold values. Thresholds determine where stoplights will indicate a "go–caution–no-go" signal. It is unusual for groups to merely accept the preset values when making a decision, rather (as explained below) they generally either set these in response to or in order to trigger stoplight events. Group 3's extraordinarily high percentage of changes to thresholds suggests a focused examination of the actual clickstream data might be in order; that Group 5 never adjusted these thresholds invites speculation as to why no changes were made.

Decision Aid Tool—Lag Sequential Analysis

We used the lag sequential analysis method described by Bakeman and Gottman (1997) to investigate dyads of clickstream sequences. The results of this analysis are presented in a longer version of this paper, which available at http://btl.usc.edu/edo/toolResults.doc.

Lag sequential analysis provided insight on how the students incorporated risk management theories into the problem-solving process. In general, the students were less reticent to manipulate utility values than probability values and, in fact, they performed this type of action significantly more often than expected. The data collected from six student groups in two different classes suggested that all the students used one of either two different strategies:

> *Strategy 1*: A global approach, in which students decide on all possible procurement options, and then set utility values and probability values independently of one another. Groups 1 and 4 displayed this type of strategy.
>
> *Strategy 2*: An option-by-option approach where students set the utility values for each event of a single procurement option, then set the probabilities for the likelihood that each event within that option might occur. The process is repeated for each subsequent procurement option. In this strategy, the action of setting utility and probability values are more closely integrated than they are in Strategy 1. Groups 2, 3, 5 and 6 displayed this type of strategy.

Student Survey

The details of the student survey results are also reported in http://btl.usc.edu/edo/toolResults.doc. Although students reported some impact of the tool on their own understanding of the RAS problem, they reported much more impact on their team. The pattern of correlations suggests that the DAT may have been a focal point around which the team could jointly explore the problem space, with decisions made explicit through the use of this tool.

Discussion and Conclusion

Rather than merely ask students what facts of knowledge they had acquired about the U.S. military’s procurement process, we asked that they actually use a Decision Aid Tool and apply the concepts of expected value and multi-attribute utility theories to evaluate procurement options for a new RAS system. In both cases, the context of the assessment was very different from other assessments that students had previously received. Unlike written tests, the DAT required the students to apply what they had learned to resolve an authentic problem. Unlike the evaluation of student presentations, the DAT attempted to uncover the underlying process that students used to reach a conclusion.

Our experience with the six groups in this study suggests that if specific evaluative criteria could be established, the DAT tool could be developed into a tool that could both teach and evaluate student strategies in near-real time. As Wolf, Bixby, Glen and Gardner (1991) suggest, such assessments can become episodes of learning.

The development of the DAT makes use of the iRides simulation language and graphics. The resulting product provides an environment in which concepts taught in the course can be applied in a tool context that can be quickly learned. However, the product does not make direct use of any of the pedagogical features of iRides. As Self (1995) has shown, simply providing an interactive environment for experimentation is not sufficient to result in timely learning. We plan to offer brief lessons in the context of a simplified DAT analysis of the "choosing a restaurant" case for use in the class on multi-attribute decision theory. This will make it possible for EDO School instructors to introduce core concepts of the tool in advance of teaching about Expected Value Theory.

References

Airasian, P. W., & Walsh, M. E. (1997). Constructivist Cautions. *Phi Delta Kappan*, **78**, 444 – 449.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: an introduction to sequential analysis*. New York: Cambridge University Press.

Bakeman, R., & Quera, V. (1995). *Analyzing interaction: sequential analysis with SDIS and GSEQ*. New York: Cambridge University Press.

Baker, E. L. (2002). Design of automated authoring systems for tests. In N. R. Council (Ed.), *Technology and assessment: thinking ahead* (pp. 79 - 89). Washington DC: National Academy Press..

Bransford, J. D., & Schwartz, D. L. (Eds.). (1999). *Rethinking transfer: a simple proposal with multiple implications* (Vol. 24). Washington DC: American Educational Research Association.

Brown, J. S., & Duguid, P. (2000). *The Social Life of Information*. Cambridge, MA: Harvard Business School Press.

Camacho, M., & Good, R. (1989). Problem solving and chemical equilibrium: Successful versus unsuccessful performance. *Journal of Research in Science Teaching, 26*, 251 - 272.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanation: how studetns studey and use examples in learning to solve problems. *Cognitive science, 13*, 145 - 182.

Chi, M. T. M., & Bassok, M. (1989). Learning from examples vis self-explanations. In L. B. Resnick (Ed.), *Knowing, learning and instruction: Essays in Honor of Robert Glaser*. Hisllsdale NJ: Lawrence Earlbaum Associates.

Cole, M. (1996). *Cultural psychology: a once and future discipline*. Cambridge MA: Harvard University Press.

Edelstein, R. A., Reid, H. M., Usatine, R., & Wilkes, M. S. (2000). A comparative study of measures to evaluate medical students' performances. *Academic medicine, 75*(8), 825 - 833.

Glaser, R., & Baxter, G. P. (2000). *Assessing active knowledge* (No. CSE Technical Report 516). Los Angeles: UCLA / CRESST.

Ge, X., & Land, S. M. (2002). The effects of question prompts and peer interactions in scaffolding students' problem-solving processes on an ill-structured task. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

deGroot, A. D. (1965) *Thought and choice in chess*. The Hauge: Mouton, 1965.

Kilonsky, D. J. (2002). Constructing knowledge in the lecture hall. *Journal of college sceince teaching, 31*(4), 246 - 251.

King, A. (1994). Guiding knowledge construction in the classrooom: effects of teaching children how to qeustion and how to explain. *American Educational Research Journal, 31*, 338 - 368.

Lesgold, A. M. (1987). Toward a theory of curriculum for use in designing intelligent instruction systems. *ERIC, ED 288 493*.

Loewen, E. R., Shaw, J. R., & Craik, F. I. M. (1990). Age differences in components of metamemory. *Experimental Aging Research, 16*, 43 - 48.

Lowyck, J., & Poysa, J. (2001). Design of collaborative learning environments. *Computers in human behavior, 17*, 507-516.

Mann, R., & Jepson, A. (1993, October 1994). *Non-accidental features in learning.* Paper presented at the AAAI Fall Symposium on machine learning in vision, Raleigh NC.

Mayer, R. E., & Wittrock, M. C. (1996). Problem Solving transfer. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47 - 62). New York: Simon & Schuster Macmillan.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational Measurement* (pp. 13–103). New York: McMillan.

Munro, A. (2003). Authoring simulation-centered learning environments with Rides and Vivids. In T. Murray, S. Blessing, & S. Ainsworth (Eds.), *Authoring tools for advanced technology learning environments*. Dordrect, The Netherlands: Kluwer Academic Publishers.

Munro, A., & Pizzini, Q. A. (1998). *VIVIDS Reference Manual*. Los Angeles: Behavioral Technology Laboratories, University of Southern California.

Munro, A., Surmon, D., Johnson, M., Pizzini, Q., & Walker, J. (1999). An open architecture for simulation-centered tutors. In S. P. Lajoie & M. Vivet, *Artificial intelligence in education: Open learning environments: New computational technologies to support learning, exploration, and collaboration* (pp. 360-367). Amsterdam: IOS Press.

Munro, A., Surmon, D., & Pizzini, Q. (in press). Teaching procedural knowledge in distance learning environments. In H. F. O'Neil, Jr. & R. S. Perez (Eds.) *Web-based learning: Theory, research, and practice*. Mahwah, NJ: Lawrence Erlbaum Associates.

Norman, D. A. (1993). *Things that make us smart: defending human attributes in the age of the machine*. New York, NY: Addison-Wesley.

Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington DC: National Academy Press.

Piaget, J. (1964). Development and learning. In R. E. Ripple & V. N. Rockcastle (Eds.), *Piaget rediscovered: report on the conference on cognitive studies and curriculum development* (pp. 7-20). Ithaca, NY: Cornell University Press.

Quellmalz, E., Schank, P., Hinojosa, T., & Padilla, C. (1999). Performance Assessment Links in Science. *Practical Assessment, Research & Evaluation, 6*(10).

Renkl, A. (2002). Worked-out examples: insructional explanations support learning by self-explanation. *Leraning and instruction, 12*, 529 - 556.

Richards, W., Feldman, J., & Jepson, A. (1992). From features to perceptual categories. In D. Hogg & R. Boyle (Eds.), *British Machine Vision Conference 1992* (pp. 99-108). Berlin: Springer-Verlag.

Richards, W., Jepson, A., & Feldman, J. (1996). Priors, perferences and categorical percepts. In D. C. Knill & W. Richards (Eds.), *Perception as baysian inference*.

Scardamalia, M., & Bereiter, C. (1993). Technologies for knowledge-building discourse. *Communications of the ACM, 36*(5), 37-41.

Self, J. (1995). *Problems with unguided learning.* Proceedings of the International Conference on Computers in Education. ICCE'95.

Shavelson, R., Gao, X., & Baxter, G. P. (1993). *Sampling variability of performance assessments* (No. CSE Technical Report 361). Los Angeles: UCLA / CRESST.

Shepard, L. A. (2000). *The role of assessment in a learning culture.* Paper presented at the Annual meeting of the American Educational Research Association, New Orleans.

Sivin-Kachala, J., & Bialo, E. R. (1998). *Report on the effectiveness of technology in schools, '90 - '97*: Software Publisher's Association.

Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: underlying principles and essential elements. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 267 - 306). Mahwah NJ: Lawrence Erlbaum Associates, Inc.

Sternberg, R. J. (1999) Intelligence as developing expertise. *Contemporary Educational Psychology*, 292-316.

Watanabe, S. (1985). *Pattern Recognition*. New York: John Wiley & Sons, Inc.

Wiggins, G. P. (1993). *Assessing Student Performance: exploring the purpose and limits of testing*. San Francisco: Jossey-Bass.

Witkin, A. P., & Tenenbaum, J. M. (1983). *What is perceptual organization for?* Paper presented at the International Joint Conference on Artificial Intelligence, Karlsruhe, West Germany.

Wolf, D.P., Bixby, J., Glen, J., & Gardner, H. (1991). To use their minds well: investigating new forms of student assessment. *Review of Research in Education, 17*(2), 31-37.