The Alberta Journal of Educational Research Vol. XLV, No. 4, Winter 1999, 392-408

Carl Frederiksen

and

Janet Donin
McGill University

Cognitive Assessment in Coached Learning Environments

An approach to cognitive assessment of problem-solving in complex computer-based tutorial environments is described. The approach is based on studies of expert tutoring and students' performance in natural tutoring situations in specific domains such as engineering and statistics. A model of expert tutors' knowledge in a domain of applied statistics was developed and used as a basis for a web-based computer coach that emulates human tutoring. Cognitive assessments are obtained from records of students' actions as they learn to apply particular components of the procedural knowledge required to solve problems in the domain with the help of the computer tutor. Learning is evaluated by studying changes in these records of performance as students practice successive problem exercises. These assessments can then be used subsequently to predict students' unassisted performance in solving post-instruction transfer problems.

Cet article décrit une approche à l'évaluation cognitive de résolution de problèmes dans un contexte complexe de tutorat assisté par ordinateurs. L'approche est fondée sur des études de tutorat par des experts et le rendement d'apprenants dans des situations naturelles de tutorat pour des domaines précis tels le génie et les statistiques. On a développé un modèle des connaissances d'experts-tuteurs en statistiques appliquées qui a servi comme base pour un tuteur informatique qui imite, sur le Web, les tuteurs humains. On obtient des évaluations cognitives à partir du record des démarches entreprises par les apprenants qui assimilent des éléments des connaissances procédurales requises dans la résolution de problèmes à l'aide du tuteur informatique. L'apprentissage est évalué en étudiant les changements dans la performance des apprenants alors qu'ils tentent la résolution de problèmes successifs. Ces évaluations peuvent ensuite servir dans la prédiction du rendement post-instructif des apprenants qui résoudront des problèmes de transfert sans aide.

There is a growing interest among educators in problem-based, collaborative, and apprenticeship approaches to instruction. This has been accompanied by an increased interest on the part of cognitive psychologists in how to model such complex instructional situations and the learning processes that occur in them. Interest in coached, collaborative, and problem-oriented modes of instruction has led to a cognitive apprenticeship model (Collins, Brown, & New-

Carl Frederiksen is a professor in the Department of Educational and Counselling Psychology. He previously held positions at the University of California, Berkeley, the National Institute of Education, and Rockefeller University. His current research specialization is in the study of discourse processing, situated cognition and learning, and learning through interactive discourse (e.g., in tutoring, problem-based learning groups, computer tutoring environments).
Janen Donin is an associate professor in the Department of Educational and Counselling Psychology and the Department of Second Language Education. Her field of specialization is discourse processing, learning, and the acquisition of information from written and oral language (in a first or second language).

man, 1989). This model has provided a rationale for reconceptualizing how students learn and how computers may be used to support these modes of learning. Computer-based learning environments currently are being designed to supplement and enhance natural problem-based learning situations by providing tools and coaching support for learners that are designed explicitly to support these kinds of learning (Derry & Lajoie, 1993). The challenge is to design tools that can support students' individual and collaborative development of complex, real-world knowledge and skill in authentic domains of knowledge and contexts of problem-solving.

As cognitive models of learning have expanded in their capacity to address complex, real-world cognition, there has been a growing recognition of the need for assessments of students' cognitive processes, learning, knowledge, and skill development that are more appropriate to these natural situations of coached, collaborative, and problem-based learning than are standard assessments of ability or achievement (Snow & Lohman, 1993). This recognition has been apparent in calls for alternative approaches to assessment (reviewed by Birenbaum, 1995). Among the characteristics that have been advocated for alternative assessment are: (a) assessment should be integrated with instruction; (b) it should be transparent, that is, it should help students learn to monitor, self-evaluate, and reflect on their own performance; (c) assessment tasks should be authentic, extended over time, meaningful, and challenging; (d) students should have access to tools, resources, and coaching support during assessment of their performance and learning; (e) assessment should be diagnostic, providing information about students' knowledge, cognitive processes, misconceptions, and errors during performance of problem-solving tasks; (f) students' knowledge and their ability to apply (transfer) their knowledge to new or novel problems should be assessed; and (g) ability to collaborate effectively with others in solving problems should be assessed. These characteristics of alternative assessment address limitations of traditional testing approaches and promote more authentic assessments of knowledge, learning, cognitive processing, and problem-solving skill in complex natural domains and contexts of task-oriented action (e.g., in higher education). However, experience with alternative assessment has raised concerns about issues of objectivity, reliability, and validity of such assessments, particularly when the situation is "high stakes" for the students being assessed. Finding objective ways to achieve more authentic cognitive assessments represents one of the great challenges for the field of measurement and evaluation.

The need for cognitively valid assessments has been recognized by researchers in the measurement community who are attempting to develop measurement models and tests that can better assess cognitive processes and knowledge structures (Frederiksen, Mislevy, & Bejar, 1993). This problem can be seen from both a cognitive perspective and a measurement perspective. From the cognitive perspective the problem is how can we use observations of students' performance in problem-based instructional environments (such as tutoring and coached instruction) to provide diagnostic information about their cognitive processes as they acquire knowledge and effective problem-solving skills through coached practice in solving problems in a domain? From the measurement perspective the problem is how can measurement models and techniques be used to provide cognitive assessments of skill and knowledge acquisition that go beyond the performance of specific tasks and permit generalizations and prediction of performance on other tasks in a domain.

Work has already begun on the first problem (Frederiksen, Glaser, Lesgold, & Shafto, 1990; Snow & Lohman, 1993), and some of this work has given high priority to the authenticity of the tasks and learning environments in which diagnostic assessment is carried out (Lesgold, Lajoie, Logan, & Eggan, 1990). Work on the second problem has typically required the construction of tasks or items that are representative samples of tasks in a domain, and the application of measurement models such as item response theory (IRT) to construct statistical models that are capable of estimating an examinee's level on a latent ability scale (or classification into a nominal or ordinal category) that pertains to the domain of tasks being sampled. Attempts are being made to construct items and IRT models that can permit some degree of cognitive assessment (Mislevy, 1993). For example, Embretson (1993) has constructed IRT models for spatial rotation tasks that decompose these tasks into subskill components.

This article is concerned with both problems. Our research initially focused on the first problem: diagnostic assessment of students' performance as they learn to solve complex problems with the help of a tutor. Currently it is addressing the second problem by applying a dynamic assessment approach that makes use of performance data that are gathered as individuals practice solving problems with the support of a coach or tutor. Measurements can be derived from such performance data, and these may be used to predict unassisted performance on new transfer problems presented following instruction. However, because psychometric generalizations typically depend on samples of tasks (i.e., problems), the question can be posed: To what extent can we predict performance on a criterion problem from dynamic assessment data gathered during instruction (Campione & Brown, 1990)? If dynamic assessments of growth in proficiency across a set of practice problems can predict criterion performance on new problems, then we may be able to claim some degree of generalizability without the need to test individuals on large samples of problems. If an individual possesses a stable and integrated domain knowledge structure and skill in applying it to perform tasks effectively in a domain, this might constitute a latent trait-like characteristic of the individual that could be assessed and that would reliably predict performance on tasks in the domain. This seems to be the implicit model for assessment in many domains of professional education.

As a first step to a dynamic and cognitive assessment of performance, we have been studying students' problem-solving and learning processes in natural tutoring situations in a domain of applied statistics. The topic is analysis of variance (ANOVA). We have found that it is possible to construct computer-based coached-learning environments that emulate many aspects of the interactive support coaches provide to students in these natural tutoring situations. We have implemented such a computer environment in the domain of applied statistics (as an ANOVA Tutor) and are currently conducting studies to assess students' learning and problem-solving while using the tutor. The ANOVA Tutor incorporates a model of the expert problem-solving knowledge that was used by an experienced tutor to coach students in how to solve problems in the domain of analysis of variance. When used in conjunction with practice data sets and appropriate statistical analysis software, the ANOVA Tutor provides a well-defined, interactive, problem-based learning environment in which students' learning and problem-solving may be assessed under realistic conditions of coached practice and instruction.

Development of a Computer Tutor in Statistics

In a previous study of students' learning in a face-to-face tutoring situation in engineering, analyses were made of changes in individual students' and the instructor's performance during coached practice in solving a sequence of problems that increased in complexity (Frederiksen, Roy, & Bédard, 1999). The changes in the tutor's explanations and scaffolding of students' problem-solving that were observed as a student gained knowledge and skill provided an almost classic example of cognitive apprenticeship. On the first problem the tutor demonstrated how to solve an example of an engineering problem (in the domain of mechanical engineering) and explained her solution procedures to each student as she solved the problem. Analysis of the tutor's discourse and problem-solving actions enabled us to develop a hierarchical model of the structure of the problem-solving procedures that the tutor modeled for the students. It also enabled us to model how the tutor explained the component procedures through her dialogue with the students. On subsequent problems, the tutor switched to a coaching role in which she asked each of the students to try to solve a new practice problem by themselves. As a student attempted to solve the problem, she provided guidance and help when she felt it was needed. Analysis of the students' problem-solving protocols and discourse interaction with the tutor in these coaching sessions revealed a learning process in which the students attempted to apply the component procedures to new problem examples (with guidance from the tutor) and obtained help from the tutor when it was needed to complete the problem. By the third problem all the students required less help from the tutor.

The results of this study of tutoring in engineering inspired the approach we are taking to cognitive assessment in the present research. This approach combines dynamic assessment of learning (i.e., providing coaching support to students with challenging problems and recording students' growing ability to solve problems without assistance) with *diagnostic assessment of students' problem-solving processes* (based on an expert model of conceptual knowledge and problem-solving procedures in a domain). Our approach is to develop a computer coach that emulates aspects of human tutoring and assessment techniques for this coaching environment that are parallel to those used to evaluate students' learning in the human tutoring situation. The engineering tutoring study demonstrated how an expert model of complex procedural domain knowledge can be constructed using knowledge-modeling and discourse-analysis tools from cognitive science to analyze the discourse and problem-solving actions of the tutor as she demonstrated and explained the problem-solving procedure. The next step was to apply these methods to analyze tutoring in the new domain (ANOVA as taught to doctoral students in educational and counseling psychology) and develop a computer coach based on an analysis of expert tutoring in the domain.

To construct the computer coach we used a program we developed called Tutor Builder to construct a database of *tutoring knowledge* from our analysis of the tutor's demonstrations and explanations of how to solve data-analysis problems in statistics. This database consists of a hierarchical data structure containing a large number of HTML files that contain information about component procedures (similar to that provided by the human tutor). While students run a statistics program (e.g., SYSTAT) to analyze a practice data set on their computers, they can run the ANOVA Tutor program concurrently on a remote server using a web browser. The students can use the browser to view and interact with a hierarchical guide to the organization of problem-solving actions. They can also view multimedia messages explaining particular steps in solving data analysis problems. In this way students can use the tutor to obtain instruction and coaching support as they practice solving data analysis problems on their computer. The computer tutor and statistics software together provide a well-defined, coached, problem-based learning environment in statistics.

Description of the Statistics Tutoring Situations

We have been studying several types of tutoring situations in statistics. These include: (a) face-to-face tutoring of individual students (in which the tutor and student shared the use of statistical software and the tutor used the software and prepared data files to demonstrate and explain how to use ANOVA to solve data analysis problems); and (b) networked one-on-one tutoring (in which the tutor introduced a student to ANOVA in the same manner as in the face-to-face condition, but communicating by means of videoconferencing software). In both of these situations doctoral students' in educational and counseling psychology shared the use of statistical software (SYSTAT) as they learned to use ANOVA to solve a series of problems consisting of data sets to be analyzed. A stack of "blackboard" representations (e.g., graphics, equations, tabular, and other displays) were provided as resources for the students as they were tutored by an experienced faculty member in ANOVA theory and methods, and in how to use SYSTAT as a tool for data analysis.

By studying tutoring in situations in which communication with the tutor was either face-to-face or by means of videoconferencing, we were attempting to bridge the gap between authentic cognitive apprenticeship situations (e.g., face-to-face tutoring) and computer environments that simulate the conditions of natural expert tutoring. Our results indicated that networked tutoring is similar to face-to-face tutoring and that the same models of expert knowledge and tutoring skill apply across all these situations. The next step is to see if a computer coach can be introduced to emulate coaching and explanation functions of a human tutor and to analyze students' learning (individual and collaborative) in these environments. Computer coaches designed in this way can offer many of the benefits of human tutoring. They can be used over a network. Moreover, they could be well suited for use in university courses as a complement to the activities of instructors. In the present context we are interested in the potential use of these environments for cognitive assessment purposes.

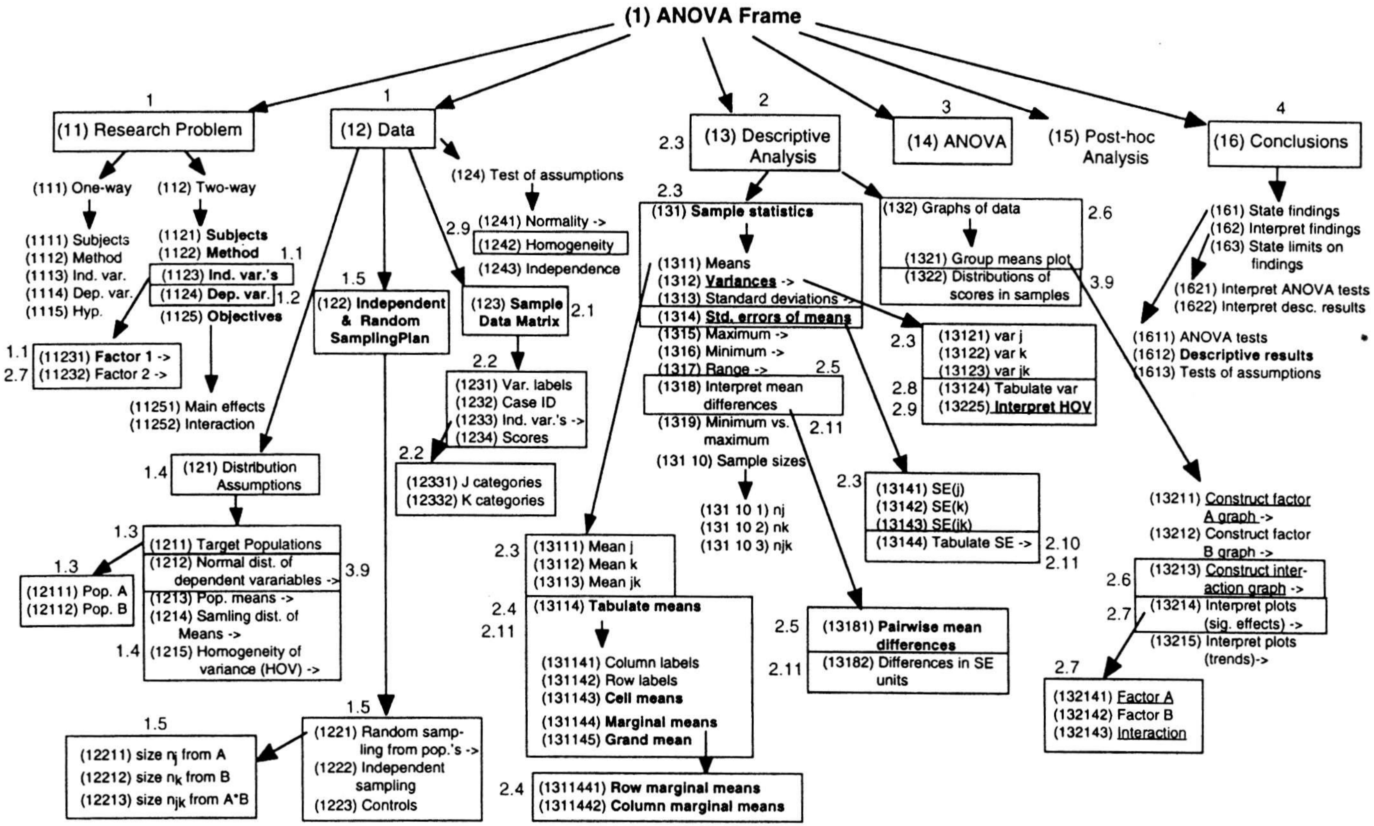

Figure 1a. Structure of a procedure frame for defining an ANOVA Tutor database.

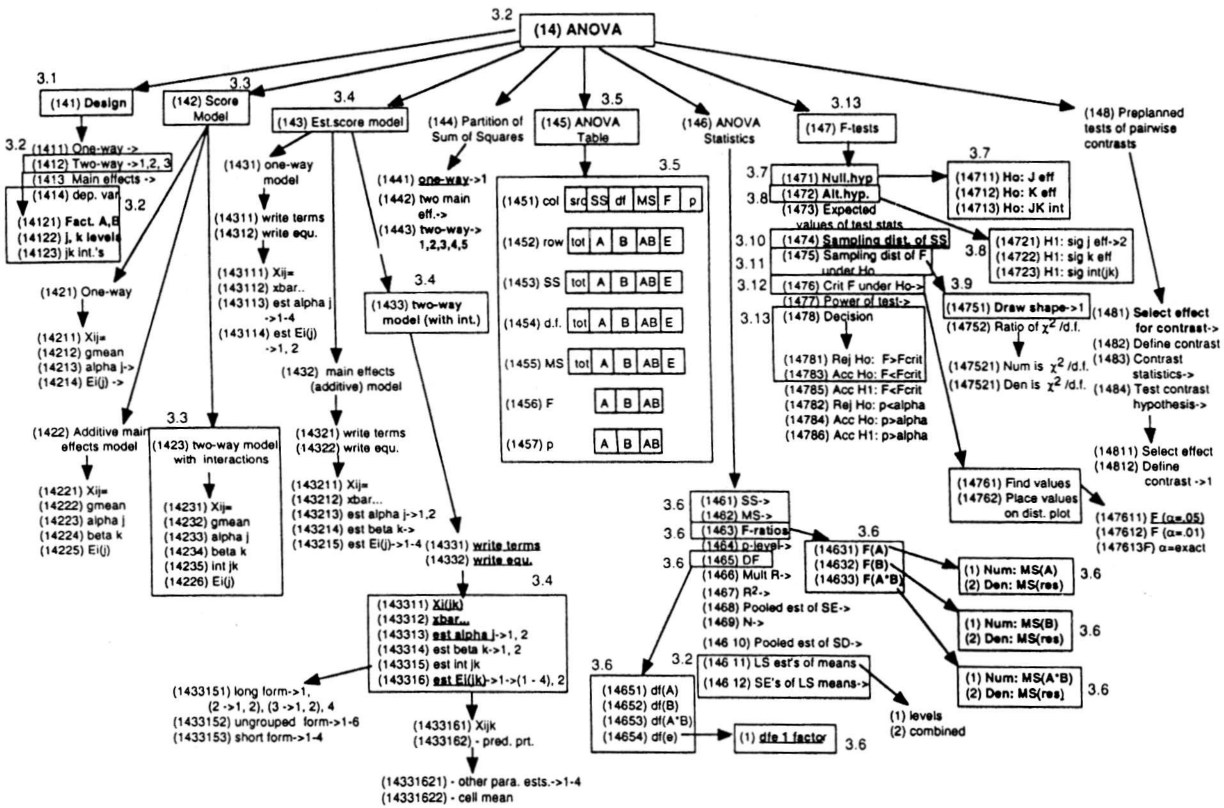

Figure 1b. Structure of a procedure frame for defining an ANOVA Tutor database.

Modeling Expert Knowledge: The Development of an ANOVA Procedure Frame

As in the previous study of tutoring in engineering, an expert model of procedural knowledge in the domain of analysis of variance was developed. This model was developed initially from a procedure frame and semantic analysis (Frederiksen, 1986) of tutorial dialogue between an experienced tutor and a novice student in a context of tutor demonstration and explanation of analysis of variance. This discourse occurred over three tutoring sessions in which the tutor modeled problem-solving procedures and provided a conceptual explanation of one- and two-way analysis of variance for a single novice student. Subsequently, data from the tutoring of other students were used to improve the model. The discourse and protocol analysis led to a relatively complete and detailed "expert model" of analysis of variance procedures.[1] This model is displayed in Figure 1.

The procedure frame represents a complex procedure by decomposing it into a hierarchy of actions and goals. At the top level, solving a data analysis problem involves six component procedures (or tasks): (a) defining the research problem, (b) specifying the data, (c) carrying out a descriptive analysis of the data, (d) performing an ANOVA with the data, (e) conducting any post-hoc analysis, and (f) drawing conclusions based on the results obtained from previous steps. Each of these main procedures is composed of subprocedures. For example, procedure (d), performing the ANOVA, is composed of eight main subprocedures to be performed: (a) specifying the research design, (b) specifying a linear model for scores on the dependent variable, (c) obtaining least squares estimates of the grand mean and all effects in the linear model, (d) partitioning the total sum of squares according to the ANOVA model, (e) preparing an ANOVA table for organizing results, (f) computing ANOVA statistics, and (g) conducting F tests. An additional procedure of conducting preplanned tests of contrasts is optional.[2]

The tutor's coverage of the analysis of variance procedure was impressive for its consistency, completeness, and explicitness. The only major topics not covered by the tutor were (a) the procedures for testing assumptions in ANOVA (procedure node 124 in Figure 1a), and (b) procedures for preplanned tests of contrasts (procedure node 148 in Figure 1b). Concepts and theory were included in the form of embedded explanations of conceptual knowledge and reasoning underlying procedures used in the analysis. The tutor's discourse provided detailed explanations of the relevant statistical theory and concepts, related this conceptual knowledge to appropriate steps in the procedure, and modeled statistical reasoning associated with applying this knowledge. The tutor's descriptions of procedures included the same kinds of information about procedures as were found for the engineering tutor. These included: *Goals* (to be attained for a current procedure); *Actions* (to be performed); *Results* (obtained from enacting the procedure); *Explanations* (several types); and *Tools* (software or other tools used to carry out a procedure and descriptions of how to use them). Explanations included: *Representations* (formulas, equations, tables, graphs); *Concepts* (statistical concepts such as mean square, *F* ratio, null hypothesis); *Theory* (statistical theory underlying a procedure); *Procedure* (the rationale for a step in the procedure); *Results* (explaining the meaning of results); and *Pragmatic Explanations* (practical information about contexts in which a procedure is applied).

What was particularly striking about the statistics tutoring situation as an example of cognitive apprenticeship was that this wealth of information about how to solve data analysis problems was embedded in contexts in which it was being used to understand problems and apply appropriate methods to solve them. Our research demonstrated that it was possible to apply cognitive modeling and discourse analysis techniques successfully in this large and complex domain and that these techniques could be used to construct a detailed model of problem-solving knowledge in the ANOVA domain. This model could be used to develop a database of tutoring knowledge for a computer tutor.

Assessment of Students' Learning

To assess students' learning from their experiences in the statistics tutoring sessions, the students participated in a third problem-solving session following the second tutoring session. This session was designed to resemble an individual coaching session, but the coach (a graduate student research assistant) provided only minimal assistance to the students. The students worked in the same environment, but were given a new data set to analyze. The coach guided each student individually through the analysis by means of a series of questions that corresponded to higher-level nodes in the procedure frame. If the student was unable to answer a question, a clarification of the question was provided, but no other form of assistance (e.g., hints, instructions, or an actual demonstration of that step in the analysis) was provided.

The mapping of assessment questions onto the procedure frame is given by the numbered boxes enclosing nodes in Figure 1. Questions are numbered consecutively within topics: Topic 1 covered the research problem and data. Topic 2 covered the descriptive analysis. Topic 3 included questions about the ANOVA itself. Topic 4 comprised questions related to the conclusions. The sequence of questions mapped onto the frame gives a *trace analysis* of how the "coach" guided the student through the procedures. An assessment of one student is indicated in the frame. Boldface node labels are used to mark procedures that were successfully executed by the student; underlined boldface nodes identify steps in the procedure for which the student made errors or displayed misconceptions. All other nodes that were the subject of questions (i.e., those enclosed in boxes) are components of the procedure that the student was unable to execute. This figure illustrates how a trace analysis (in an assessment or coaching situation) can identify procedures that have been learned, the location of misconceptions and errors, and procedures that the a student cannot execute without coaching support. Such an assessment keeps track of the extent and kinds of coaching support that a student may need for particular nodes (i.e., procedures) that have not yet been mastered.

The expert procedure frame provides a network model of the procedural knowledge that is communicated to the student and that is needed to understand and solve problems in the domain, and the trace analysis reveals how the tutoring session was organized to guide the students' problem-solving.[3] If a computer coach could be developed to provide guidance and coaching support for students' problem-solving that are similar to that of a human coach, it ought to be possible to adopt a similar approach to diagnostic assessment in the computer environment. In coaching, the trace analysis is primarily a record of the sequence in which the student applied specific procedures to solve the problem. Thus it provides an assessment of the student's ability to apply the procedures systematically to solve a particular problem. Our ANOVA tutor is designed to emulate the tutor's guidance (by means of a map of the hierarchical procedure frame) and the tutor's explanations of component tasks (by means of buttons providing access to different kinds of information about the procedure).

Tutor Builder: Constructing a Database of Tutoring Knowledge

The Tutor Builder software is currently being used to develop a database of tutoring knowledge that is structured according to the expert procedure frame that was constructed in our study of tutoring in statistics. The database is organized in terms of the procedure hierarchy, and associated with each node (i.e., component procedure) there are *semantic fields* that contain text and graphic information about the procedure. (This information is based on semantic information that was expressed by the tutor through his contributions to the tutorial dialogue.) Semantic fields provide two kinds of assistance to the student: Instruction and Coaching Assistance. *Instruction* provides several kinds of semantic description of component procedures: Goal Descriptions, Problem State Descriptions, Action Descriptions, Tool Instructions, Theory Explanations, Conditions (necessary for carrying out the step in the procedure), and Result Descriptions. *Coaching Assistance* is provided in the form of Questions, Clarifications, and Hints. As an example, consider the Tutor Window presented in Figure 2. The panel at the lower left of the Window provides the student with a "Procedure Map" of the hierarchical structure of the procedure (like a site map on the internet). The student can select any node (i.e., component procedure) from the map, and then "request" coaching assistance or instruction pertaining to the selected procedure by selecting a type of assistance (from the panel immediately above the Procedure Map). For example, types of instruction include a statement of the Goal of carrying out an ANOVA, a description of the kinds of Results obtained from doing an ANOVA, or a general introduction to the relevant Theory. *Coaching* is provided at several different levels and the student can select the level of assistance he or she desires: Questions, Clarifications, or Hints (there are up to three levels of hints available). Once a type of instruction or coaching assistance has been selected, the tutor displays the relevant text or graphic information in the right panel of the window. The buttons in this panel provide access to other types of instruction or levels of coaching. It is possible to provide graphics, sound (and even movie clips) to accompany these windows. Glossary items are highlighted in the presented text, and an alphabetized listing of all glossary items is always available by means of a pop-up selection window.

An enhanced version of the tutor (which is under development) will include facilities for students to (a) submit their work and (b) perform a guided self-evaluation by comparing their work with the "tutor's results." After students have completed a step (at any level in the hierarchy), they will be able to submit their work (in the form of a text file), and ask to view the tutor's results at this step. This opens a Result Window containing two kinds of *problem-*

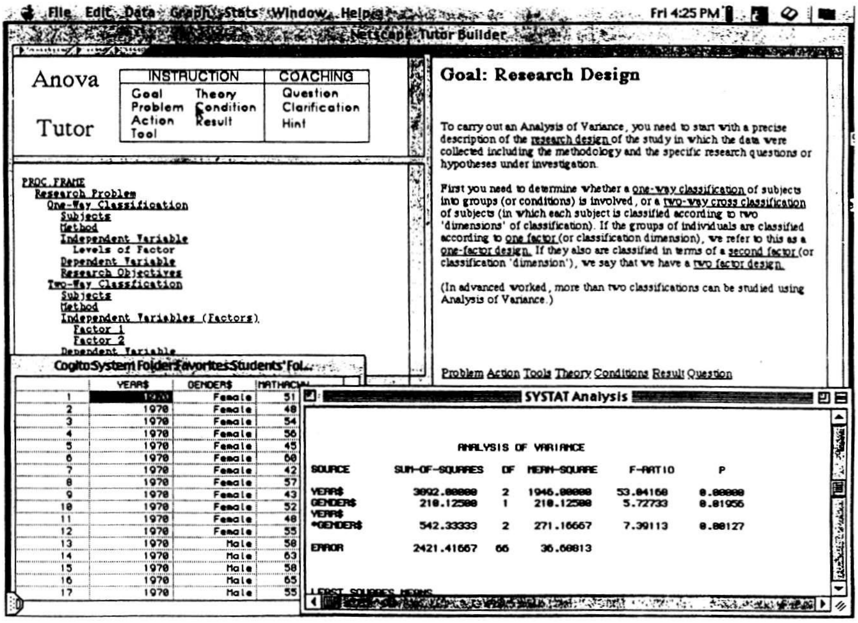

Figure 2. The ANOVA Tutor environment: Screen shot of the computer desktop with the ANOVA Tutor running (in a browser window) and with Data Editor and Analysis Windows displayed by concurrently running the statistics application SYSTAT.

specific information: first, a correct result corresponding to that step in the procedure for the problem, and second, a checklist of items required in a successful result and a list of common errors. Students will be able to compare their solutions at that step with the correct result and check those features that are present in their solutions. A student will be guided to relevant instruction based on this self-evaluation. This feature is designed to develop skill in self-monitoring and evaluation of problem-solving steps.

As shown in Figure 2, the computer coach is run concurrently with the statistical analysis software. The student is presented with a prepared data set and a description of the data, how the data were obtained, and the purposes of the study. The problem-solving task involves planning and executing a complete analysis using the statistical software, and submitting a report containing results obtained at each step (the report is organized as the sequence of student responses to tutor-supplied questions). The student will be able to cut and paste information from any tutoring window, or from the statistical environment (e.g., results from the output window) into the template, and enter information using text editing functions. This environment closely parallels natural problem-based classroom and tutoring situations. The students work on problem exercises while using the tutor. The goal is to learn to produce correct analyses independently and produce accurate and complete reports by practicing these exercises (with the help of the tutor).

Cognitive Assessment in the Coached Learning Environment

This computer coach provides an interesting environment for cognitive assessment that is authentic in the sense that it has been designed to reflect characteristics of expert tutoring and problem-solving in the domain. The authenticity of the environment can be verified by comparing it with various conditions of face-to-face and networked tutoring. If students' learning processes are found to be similar across these situations, the computer tutor could be considered to have a high degree of authenticity. A review of the characteristics often recommended by advocates of alternative approaches to assessment reveals that many of these are met, and so there is a reasonable likelihood that our coached learning environment is at least a candidate for achieving a controlled environment for cognitive assessment. In the remainder of this article we briefly examine the types of "responses" that are obtained from the students in this environment, how they can provide information related to the goals of cognitive assessment, and finally the types of statistical measurement models that appear to be needed. A thorough analysis and development of appropriate measurement models is a large task that remains to be done.

Briefly, our assessment goals are: (a) dynamic assessment of students' proficiency in problem-solving and the extent and quality of their domain knowledge; (b) diagnostic assessment of students' problem-solving: identify correct application of component procedures, location and kinds of errors, types and levels of coaching required, and pattern of application of procedures; (c) assessment of student learning: identify changes over practice problems; (d) prediction of unassisted performance on "transfer" problems; and (e) extension of assessment to collaborative learning. Our task is to specify how measures or indicators relevant to these goals can be obtained from the responses of students while working on practice problems in the coached learning environment. The responses students can make for each "item" (i.e., component procedure) are identified in Figure 3.

At any location (i.e., node corresponding to a component procedure) in the procedure hierarchy, several kinds of student response can occur. Successful completion of the subtask corresponding to a superordinate node requires successful execution of all subordinate tasks plus completion of a superordinate integrative task. For example, writing the model equation involves writing expressions for all the components of the equation plus arranging them into a single equation. For any node three types of information are recorded: (a) the student's self evaluation of his or her solution at that step in the procedure, (b) the level of coaching assistance used by the student to produce a correct solution at this step, and (c) the type of instructional assistance the student obtained from the tutor. (Note that the evaluation of a student's independent work could be done by the instructor using the tutor to "score" the student's work.) Response categories for the first two of these would be considered to be ordered: evaluations range from item correct to item failed (errors only) or not attempted, and levels of assistance range from no assistance, to questions designed to elicit appropriate knowledge, to hints that provide partial knowledge, to full instructions (each of these could have corresponding levels of detail). Response categories corresponding to types of assistance are not expected to be ordered; rather, they reflect students' preferences for different

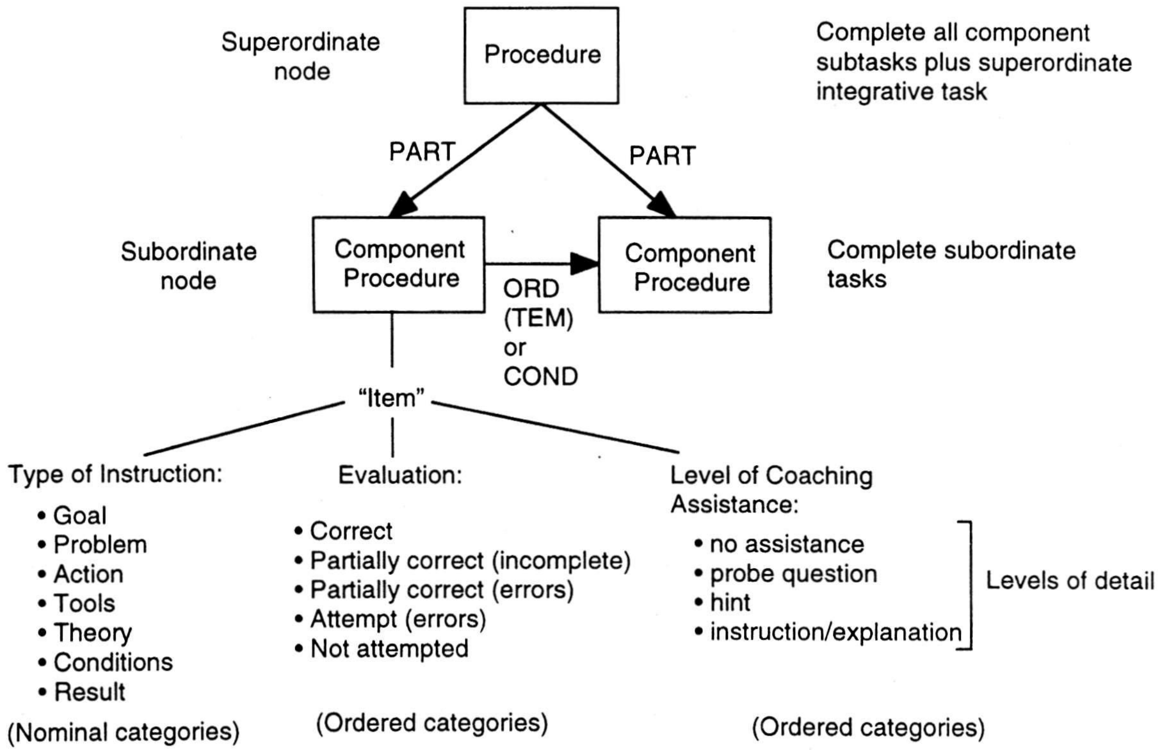

Figure 3. Response options in the ANOVA Tutor environment: Options made available by the ANOVA Tutor for each node in the ANOVA procedure frame.

types of information about procedures. The question we need to address is how can we use these responses to obtain information relevant to our assessment goals.

Domain Knowledge and Problem-solving Proficiency

Knowledge and proficiency are associated with major components (high-level components) of the problem-solving procedures. The high-level components for statistical design and data analysis problems include topics such as (a) specifying the research design, (b) specifying the sampling plan, (c) specifying and testing distribution assumptions for the data, (d) preparing the data for analysis, (e) carrying out descriptive analysis of the data, and (f) carrying out the analysis of variance. In each of these high-level components, subcomponent procedures are specified. For example, the subcomponents of (f) the analysis of variance are: (a) specifying the appropriate ANOVA design, (b) writing the linear score model, (c) obtaining estimates of parameters in the model, (d) partitioning the sum of squares, (e) computing ANOVA statistics, (f) conducting significance tests, (g) constructing the ANOVA table, and (h) carrying out tests of preplanned contrasts. Because each of these decomposes hierarchically into components, proficiency is associated with successful execution of the components and integration of components at each level into a solution of the more general problem.

Two kinds of measurement models might be considered: hierarchical models that explicitly incorporate the hierarchy into the measurement model (Embretson, 1993; Gitomer & Rock, 1993) and latent trait models that scale *items* (corresponding to the procedure components) on a single dimension or *proficiency scale.* A complicating factor is that measurement based on these items may be based on ordered categories of response (levels of correctness or assistance), suggesting some kind of partial-credit model (Masters, 1982; Masters & Mislevy, 1993) or assignment to purely categorical latent classes (e.g., based on categories corresponding to types of assistance, Yamamoto & Gitomer, 1993). Test theory models of this kind may be applicable to response measures obtained in the coached learning environments, and would justify assessments at the level of the "large" procedure components that are involved in data analysis using ANOVA models and methods. The function of such assessments would be to certify that a student had reached a particular level of proficiency and knowledge with respect to a given component procedure.

Diagnostic Assessment of Component Knowledge and Problem-solving Procedures

Diagnostic assessment makes a less stringent demand on statistical measurement models if the primary goal is to provide the student or instructor with diagnostic information about where the student is having difficulty and what kinds of knowledge are lacking. Such information is provided by the trace of student responses using the tutor and evaluations of the students' work. Reliability of such diagnostic assessments is less of a concern, but consistency of errors in component tasks or need for assistance would indicate real misunderstandings or lack of knowledge. Properties of individual items (components) resulting from item analyses performed to construct proficiency measures would provide evidence of how significant difficulties at the level of specific items are as contributors to proficiency on major task components. In addition to diagnosing where difficulties occur, diagnostic assessment would include trace analysis of particular patterns and sequences in a student's application of procedures. It may be possible to establish expert-like patterns and novice patterns to use as a basis for interpreting students' sequence of problem-solving operations (Lesgold et al., 1990). Finally, the quality of student conceptual understanding and explanations might be assessed at various steps in problem-solving to evaluate students' knowledge. This could be done through reports, questions, or other tasks designed explicitly to obtain such information from the students.

Assessment of Change with Practice

The assessment of change can occur both at the level of domain knowledge and proficiency and at the level of individual component procedures and the sequence in which they are applied (diagnostic level). The principal question is to what extent is a student's performance gradually approximating that of an expert? Answering this question will involve interesting analyses of changes in the traces of students' responses while using the tutors. In addition to the pattern of performance finally attained, changes in students' performance over problem exercises provides information about how an individual progresses and any learning difficulties that may have occurred. Such data may be valuable in helping students develop effective learning strategies for complex domains such as statistics.

Prediction of Performance on Criterion Problem-Solving Tasks

A straightforward approach to prediction of performance would be to assess proficiency with main component procedures while using the tutor when solving practice problems. Of particular interest would be the level of proficiency attained for each procedure and the difficulty level of the practice problems attempted. The same assessment criteria could be applied to a post-practice near-transfer assessment problem, and then the tutor performance could be used to predict final performance. If tutor performance turns out to be a good predictor of performance on the criterion problems, then the use of a coached-learning environment for assessment might be recommended. In addition, a close analysis of students for whom predictions are poorest might provide interesting information about factors responsible for difficulty in making the transition from assisted performance to independent competence in a domain. This issue corresponds to one of the principal arguments for a cognitive apprenticeship approach in which assistance provided by a coach is gradually reduced to enable students to develop autonomous skill in a domain.

Extension to Collaborative Learning Situations

Coached learning environments can be used effectively by pairs of students or small groups of students, with the added benefits of motivation and development of skill in collaborative problem-solving. As an environment for cognitive assessment, the trace of students using the environment would reflect the knowledge of each and their joint problem-solving behavior. When used this way, it would seem essential that students also attempt problems individually (also with the help of the tutor) and that they be assessed independently in order to develop independent competence and knowledge in the domain. Data obtained in such studies would bear on many of the learning issues underlying collaborative approaches to instruction, especially on the relationship between collaborative competence and individual competence and knowledge in a domain.

In summary, coached learning environments of the type we describe in this article appear to offer potential as environments for assessment. As assessment environments they can provide objective and cognitively valid indicators of performance. Measurement models can be developed for students' responses in these environments to enable assessment of gains in knowledge and proficiency, and these can be validated as predictors of performance on criterion problem-solving tasks. Moreover, the performance measures themselves can be used to provide instructors (and students) with diagnostic information about students' understanding and problem-solving, including their errors and misconceptions. In addition, coached-learning environments can provide a high degree of authenticity as natural environments for collaborative and problem-based learning. As such they should provide valuable contexts for coached practice in problem-solving that can extend and supplement classroom instruction. If future research can show that this approach to assessment can be implemented in conjunction with appropriate cognitive and measurement models to create reliable assessments with high predictive validity, then it may offer one solution to the problem of assessment that is objective, authentic, and cognitively valid.

Notes

1. The analysis of the tutorial discourse was carried out using a computer-aided analysis tool (NU•DIST) to build the frame. The complete analysis of tutorial dialogue that led to the development of the procedure frame is stored in a NU•DIST database. This database documents how each text unit (and associated actions) in the tutorial dialogue was matched to procedures in the expert model.
2. Note that Figures 1a and 1b do not include all of the branching procedures: branching subprocedures are collapsed into lists in the figures and if they are followed by arrows, they branch further.
3. In coached tutoring sessions (such as those analyzed in our engineering tutoring study), the tutor and student participated more equally in the dialogue with the tutor occasionally guiding the student to the next high-level procedure, but with the student initiating most procedures and related dialogue exchanges.

Acknowledgments

This research was supported by the Telelearning NCE, funded by the Government of Canada. The development and testing of the statistics tutor are being supported by a grant from the Office of Learning Technologies, Human Resources Development Canada. The engineering tutoring study was supported by the National Research Council and BNH Expert Engineering Software, Inc. St. Laurent, QC. Lorraine Meilleur and Marguerite Roy assisted in the research, and Michel Décary wrote the Tutor Builder software. A special thanks to Robert Bracewell for participating as the statistics tutor in these studies.

References

Birenbaum, M. (1995). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum & F.J.R.C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes, and prior knowledge* (pp. 3-29). London: Kluwer.

Campione, J.C., & Brown, A.L. (1990). Guided learning and transfer: Implications for approaches to assessment. In N. Frederiksen, Glaser, R., Lesgold, A., & Shafto, M. (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 141-172). Hillsdale, NJ: Erlbaum.

Collins, A., Brown, J.S., & Newman, D. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L. Resnick (Ed.), *Knowing, learning and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Erlbaum.

Derry, S.J., & Lajoie, S. (1993). A middle camp for (un)intelligent instructional computing: An introduction. In S. Lajoie & S.J. Derry (Eds.), *Computers as cognitive tools* (pp. 1-11). Hillsdale, NJ: Erlbaum.

Embretson, (1993). Psychometric models for learning and cognitive processes. In N. Frederiksen, R.J. Mislevy, & I.I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 125-150). Hillsdale, NJ: Erlbaum.

Frederiksen, C.H. (1986). Cognitive models and discourse analysis. In C.R. Cooper & S. Greenbaum (Eds.), *Studying writing: Linguistic approaches* (pp. 227-267). Beverly Hills, CA: Sage.

Frederiksen, C.H., & Roy, M. (1999). *Tutors' use of discourse to model and explain problem-solving procedures in statistics*. Unpublished manuscript.

Frederiksen, C.H., Roy, M., & Bédard, D. (1999). *Expert tutoring and learning: Analysis of tutorial dialogue in engineering*. Unpublished manuscript.

Frederiksen, N., Glaser, R., Lesgold, A., & Shafto, M. (Eds.). (1990). *Diagnostic monitoring of skill and knowledge acquisition*. Hillsdale, NJ: Erlbaum.

Frederiksen, N., Mislevy, R.J., & Bejar, I.I. (Eds.). (1993). *Test theory for a new generation of tests*. Hillsdale, NJ: Erlbaum.

Gitomer, D.H., & Rock, D. (1993). Addressing process variables in test analysis. In N. Frederiksen, R.J. Mislevy, & I.I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 243-268). Hillsdale, NJ: Erlbaum.

Lesgold, A., Lajoie, S., Logan, D., & Eggan, G. (1990). Applying cognitive task analysis and research methods to assessment. In N. Frederiksen, Glaser, R., Lesgold, A., & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 325-350). Hillsdale, NJ: Erlbaum.

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Masters, G.N., & Mislevy, R.J. (1993). New views of student learning: Implications for educational measurement. In N. Frederiksen, R.J. Mislevy, & I.I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 219-224). Hillsdale, NJ: Erlbaum.

Mislevy, R.J. (1993). Foundations of a new test theory. In N. Frederiksen, R.J. Mislevy, & I.I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 19-40). Hillsdale, NJ: Erlbaum.

Snow, R.E. & Lohman, D.F. (1993). Cognitive psychology, new test design, and new test theory: An introduction. In N. Frederiksen, R.J. Mislevy, & I.I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 1-18). Hillsdale, NJ: Erlbaum.

Yamamoto, K. & Gitomer, D.H. (1993). Appliction of a HYBRID model to a test of cognitive skill representation. In N. Frederiksen, R.J. Mislevy, & I.I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 275-296). Hillsdale, NJ: Erlbaum.