THE WEATHER LAB:
AN INSTRUCTION-BASED ASSESSMENT TOOL
BUILT FROM A KNOWLEDGE-BASED SYSTEM

David Mioduser

School of Education
Tel-Aviv University
Ramat-Aviv, 69978, Israel
Fax #: 972-3-6407752
e-mail: miodu@ccsg.tau.ac.il

Richard L. Venezky

Department of Educational Studies
University of Delaware
Newark, DE, 19716

Brian Gong

Learning Result Services
Kentucky Department of Education
Frankfort, KY 40601

Final and edited version to be published in:
Journal of Computers in Mathematics and Science Teaching

RUNNING HEAD: WEATHER LAB

Abstract

The Weather Lab is a computer-based tool for assessing a student's knowledge and understanding of weather phenomena. The student is involved in generating weather forecasts or in manipulating weather components (e.g., temperature or pressure measures) affecting the final formulation of a forecast. The Weather Lab is conceived as an instruction-based assessment tool, implying that:

Function: It is part of the instruction, and not a post-instruction testing device. It is aimed at informing instruction and triggering student reflections on her or his learning.

Focus: It is focused on the student's knowledge at different levels of ability (e.g., knowledge of facts, understanding of relationships, manipulation of weather variables, forecasting.)

Targets for the assessment information: The collected information is used for generating feedback loops within the system, between the system and the student, and between the system and the teacher, affecting actual instructional decision-making at these three different levels.

Assessment strategy and methods: The student/system interaction proceeds as a case-based problem solving dialog. The student may explore changes in weather variables affecting weather outcomes or generate forecasts for particular weather situations. Each interaction is assessed and evaluation information is delivered to the above mentioned targets.

Use of technology: The accumulated experience in tutoring systems development supports the building of sophisticated procedures aimed at representing and using subject-matter expertise knowledge, modeling the student performance, and generating dialog situations.

The rationale for the development of the system, an overview of its architecture, and our research agenda are presented in this paper.

The Weather Lab: An Instruction-based Assessment Tool

Built from a Knowledge-based system

This paper describes a computer-based system whose development reflects theoretical and practical issues in the design of instruction-based assessment tools. The role and character of assessment within the instructional process has been a matter of recent debate (Cognition and Technology Group at Vanderbilt, 1996; Freedle, 1990; Glaser, 1986; Harris & Longstreet, 1990; Snow, 1988). Emerging from this debate is the conviction that teaching and assessment should be planned and implemented as integral events and that in the future "learning assessments will not provide merely a score, a Label, a grade level, or a percentile. Rather we will have also 'instructional scoring' that indicates to the student and assists the teacher's judgment in making apparent the requirements for increasing competence" (Glaser, 1986, p. 56). The proposed changes in the character of this instructionally relevant assessment can be related to five main aspects: function, focus, strategies and methods, target for assessment information, and use of technology.

<u>Function</u>. Formal assessment activities are currently implemented mostly for policy and accountability purposes. The most frequent reasons for assessment administration are to summarize achievement, make selection, grouping, or placement decisions, report progress to different levels in the educational system, or evaluate teaching and programs (Wang, 1988). The proposed changes in function locate assessment within the teaching-learning process, its main purpose being to diagnose the student's state of knowledge, to deliver this information to the student and the teacher throughout the instruction (and not only at the end), to support the student's reflection on her own learning, and to support microadaptation (within an activity or a class) and macroadaptation (through the different units and topics) of the instruction (Ager, 1990; Lesgold, Ivill-Friel, & Bonar, 1989; Lipson, 1988; Snow, 1988).

Focus. Associated with changes in function is the re-definition of the focus or object of the assessment. This re-definition consists of broadening the range of knowledge levels addressed by the assessment from general abilities or relative rankings, to cover the whole spectrum of expected learning outcomes, ranging from factual knowledge to high level conceptual understanding and skillful performance. Within this spectrum, examples of potential focal points for assessment are the student's prior knowledge and theories, misconceptions, problem template recognition ability, strategies of performance and development of competence processes (Brown et al., 1993; Glaser, 1986; Lipson, 1988; Gelman & Greeno, 1989).

Strategies and methods. A third aspect to be revised are the strategies and methods used for assessment. Most of the assessment instruments currently in use, both at the level of large-scale standardized tests or at the level of local teacher-generated-tests, rely heavily on the multiple choice item format. If the assessment is to guide instructional decision making and is to address different levels of the student's knowledge, then the repertoire of assessment methods has to be enriched. One new direction uses classroom performance as a substantial source of assessment data (Bachor, 1990; Lock, 1990; Wang, 1988). Another direction promotes changes in test item formats. Such changes in format range from variants in format of multiple-choice items to creating problem solving situations where the student has to apply learned knowledge and skills, or simulation environments where the student has to create and run a model of a portion of reality. These case-based or problem solving tasks, far from the usual sterility of many testing items, resemble the real situation to which the learned concepts and skills are related. The student is required to integrate different types of knowledge, skills and strategies, in order to design and implement appropriate courses of action (Treffinger, Isaksen, & Dorval, 1994; Ward, 1986).

Still another aspect for reform is the overall design of the assessment plan. Most current standardized testing instruments are designed as collections of (mostly) independent test items. For adaptive instruction a more coherent approach is required. This can be achieved, for example, by creating sequences of tasks dealing with different aspects of the student's knowledge, or tracing the gradual development of basic concepts and skills and their successive integration into high level performance. The result should be the logical orchestration of tasks, within an instructional session as well as among stages in the instructional plan (Glaser, 1986).

Target. The recognition of the student and the teacher as potential targets or recipients of useful assessment information has practical implications not only for the design of the assessment plan, but for the design of the procedures by which the assessment results are analyzed and reported. For the information to inform the student's decisions or reflection about learning, or to guide the teacher's instructional decisions, it has to come on time and not (as the current use is) after instruction is completed. It should also come through the instructional process and in a format that can be handled easily by students and teachers. Addressing the assessment information to the student and the teacher may affect essential aspects the teaching/learning process (Gong, Venezky, & Mioduser, 1992).

Use of Technology. The last area where changes in assessment are desired is in the use of updated information technology. Although significant use of computer based tools is being made for the manipulation, implementation and analysis of large scale standardized testing (Wainer et al., 1990), few such tools exist for the student-centered and instruction-based forms of assessment mentioned in the previous paragraphs. Interesting examples of on-line feedback guiding performance can be found in some intelligent tutoring systems (Polson & Richardson, 1988; Wenger, 1987). One example is a computer-based coach that assesses the student's moves in a game-like environment and gives advice or useful information for guiding her further decisions (Goldstein, 1982). Another example is the on-line generation of test items during an instructional session according to the current knowledge state of the student as represented in the system's student model (Lesgold et al., 1989). The technological infrastructure is available to develop assessment environments that are of varied types, interactive in character, empowered with on-line analysis and report features, and oriented towards adaptive instruction. Reasonable computing power and sophisticated audio-visual features are accessible now in relatively low-cost computers. The appropriate use of this infrastructure, however, for the aims discussed here, remains a challenge (Deringer, 1986; Wang, 1988).

These proposed changes in function, focus, methods, targets and means of the assessment define the conceptual framework within which new tools and systems could be designed and implemented. What features are required for those tools and systems to embody the proposed changes? In the last decade we have witnessed the growth of a promising research and development field, dealing with the creation of sophisticated computer-based instructional systems. Under the generic denomination of intelligent computer-aided instruction (Carbonell, 1970), intelligent tutoring systems (Sleeman & Brown, 1982), or knowledge-based systems (Davis & Lennat, 1982), a broad variety of systems have been developed for tutoring, training, coaching and supporting decision making processes.

Substantial work has been done on modeling of student performances, conceptions and misconceptions as issues in themselves (Gentner & Stevens, 1983), and as part of a tutoring system (Burton, 1982; Sleeman, 1982). But in the implementation of similar strategies for the explicit generation of assessment information supporting the student's and teacher's instructional decision-making, several central questions deserve more attention than is found in the framework of existing development projects. The need for instruction-based assessment strategies to improve (and even ensure) adaptive instruction (Gong, Venezky, & Mioduser, 1992) motivates our attempt to develop a knowledge-based system as an instruction-based assessment tool.

The Weather Lab

The Weather Lab is a computer-based tool for assessing a student's knowledge and understanding of weather phenomena. The student is involved in generating weather forecasts or in manipulating weather components (e.g., temperature or pressure measures) affecting the final formulation of a forecast. Although the complete structure and features of the system will be presented later, we want to introduce here the main principles guiding its development. The Weather Lab is conceived as an instruction-based assessment tool, meaning that:

a. Function: It is part of the instruction, and not a post-instruction testing device. It is aimed at informing instruction and triggering student reflections on her learning by providing assessment information.

b. Focus: It is focused on the student's knowledge state at different levels of ability (e.g., knowledge of facts, understanding of relationships, manipulation of weather variables, forecasting on the basis of observed weather data), consistent with the types of knowledge and ability that are amenable to classroom instruction.

c. Targets for the assessment information: The collected information is used for generating feedback loops that affect actual instructional-decision-making at three different levels: within the system (what's next within the current working session as considered by the system), between the system and the student (what's next within the current working session as considered by the student), and between the system and the teacher, (what's next in the instructional plan as considered by the student and the teacher).

d. Assessment strategy and methods: The student/system interaction proceeds as a case-based problem solving dialog. The student is allowed to build a particular weather situation, to generate forecasts based on that situation, and to explore the way changes in weather variables affect weather outcomes. Each interaction is assessed and evaluative information is delivered to the above mentioned targets. The overall session plan proceeds from the initial consideration of factual data (e.g., degree of correspondence between a temperature entered by the student and the average range for a particular area in a particular month) up to the assessment of high-level conceptual performance (e.g., the student's ability to manipulate a multiple-variable model of an unusual weather situation).

e. Use of technology: Technology-related features are present in the system at different levels. In its surface they enable the building of a user-friendly and esthetic working environment. In a deeper level the accumulated experience on tutoring systems development supports the building of sophisticated procedures aimed at, for example, representing and using subject-matter expertise knowledge, modeling the student performance, and generating dialog situations.

Before we discuss specific features, we will briefly describe the Weather Lab's working environment. There are two main working modes of the system (Shown in Figure 1). The first allows the student to enter observed weather data related to temperature, wind, sky conditions, and air pressure. The student activates the instruments (e.g., thermometer or wind knob) and scales (e.g., Beaufort wind scale or a type-of-clouds table) shown in the screen to define the data on which the following work will be based. The specific data determined for each weather component is shown in the upper part of the screen. At any time during the session data items can be modified.

Insert Figure 1 about here

The second working mode is the environment where the instruction-based assessment dialog takes place. As shown in Figure 2, the data set defined in the input mode and registered in the upper part of the screen remains visible. The rest of the working space is composed of three areas. The main area is the dialog window, where most of the student/computer interactions occur. To the left of this window appears "Flake", the computer-based partner for the dialog. Clicking on Flake will always activate the next component of the session plan, according to criteria to be explained later.

Above the dialog window appears the "Your forecast" area, where the student enters her forecast. Below the dialog window is the computer forecast area, where the computer generated forecast will appear. Clicking on the "rainy-computer" button appearing on the left of this area activates the query function that allows the student to consult the system about weather facts and phenomena.

In general terms, a working session begins with the student entering a complete or partial set of weather data, according to the objectives and plan for the session. Then she enters the dialog mode where the session proceeds according to the student's interactions with the system. The student can activate at any time the different options, modifying the input data, modifying the forecast, resuming the dialog with "Flake", or consulting the system's knowledge base according to her decisions. At the end of the session a summary report is generated and communicated to the student and the teacher.

Insert Figure 2 about here

This preliminary description of the system was presented at this point to give the reader a frame of reference to attach the features and functions that are described in the rest of this paper. In the next section, excerpts from a student/computer dialog protocol are analyzed, as a way to present some of the systems features (e.g., kinds of expertise, assessment strategies). Then we will briefly elaborate on the system's design rationale, followed a section on its structure and components. In the last section we present our implementation and research agenda.

A student/system dialog

Consider the following excerpts (Figure 3) from a protocol of a student's interaction with the Weather Lab system (the student's inputs 's:' appear in bold style, the student's "clicking", selecting options or entering data in italics, and the computer outputs 'c:' in plain style).

Insert Figure 3 about here

In this passage, the system assesses the data entered by the student. Even in this short initial dialog excerpt, several features of the system are unveiled:

(a) The system is responsive to actual data, entered (and eventually changed) by the student. For example, at the beginning of the session the student sets the values of up to seven weather components (e.g., temperature, wind direction, barometric pressure). The system evaluates the actual input combinations constructed by the student (millions of possible combinations do exist. About 110,000 are relevant combinations to be considered). The system uses an inference engine, rather than "hard-wired" answers to pre-determined test items.

(b) The system "knows about" weather phenomena at two different levels. At the "single-facts" level, the system knows such things as the range of reasonable temperatures at different locations on the earth. The other, more sophisticated knowledge level involves the ability to evaluate arbitrary combinations of data. This allows the system to assess unlikely or inconsistent data, such as an unlikely temperature for a given area or period of the year, or some unlikely combination of air pressure and precipitation.

(c) The system generates and communicates its response to the student. The system's response in excerpt 1 is an example of a local assessment loop. One component --the temperature data to be used in generating the forecast-- was assessed by the system to be unlikely. The information is given to the student, leaving open two alternatives: to modify the data or to leave it as it is. In the later case the system asks the student to be aware of this irregularity, while continuing with the forecast formulation process.

(d) An important "behind the stage" feature of the system is implicit in this last item. The system's evaluation of the data entered (and potentially changed) by the student is kept by the system, to be used when required during the working session. In other words, an essential feature for an instruction-based assessment tool is the ability to keep systematic records about relevant parameters of student performance.

The following additional features of the system's subject matter expertise are illustrated in excerpt 2 (Figure 4):

Insert Figure 4 about here

(e) In addition to the knowledge about weather facts and components already mentioned, the system has knowledge about how to formulate a forecast based on the current data entered by the student. Such expertise on the different aspects of weather description and prediction is essential, since our objective is to assess the student's knowledge-state by means of exploratory and simulative instructional situations, and not against a pre-determined table of correct and incorrect answers.

(f) This passage also reveals a practical aspect of the tool: its interface features "clicking", dialog areas, buttons, and so forth. Such a graphical, user-friendly interface is essential from the real-life/school implementation perspective.

The third excerpt (Figure 5) unveils a particularly complex issue: the assessment of the student's conceptual understanding, her ability to manipulate variables, and to identify causal relationships between these variables and observable or expected weather phenomena.

Insert Figure 5 about here

(g) The system compares the student's forecast with the computer generated forecast. This is another facet of the system's expertise, focusing now on assessment strategies. The system must know when and how to assess the student's interactions. The input data assessment exemplified in the first excerpt was one example of this expertise, focusing on the "knowledge-of-facts" level. The third excerpt shows a very different knowledge-level focus, and a different assessment strategy. The target knowledge level is now the set of relationships the student has considered between given weather components and the way they affect the forecast. The student is asked to reflect on and describe the factors she has taken into account for including a given weather outcome (e.g., "little change") in her forecast.

As a result of the comparison between the student's forecast and the system's forecast components included by the student but not by the computer, and vice-versa, are identified. For the first case, rather than immediately classifying the student's excessive components as incorrect, the strategy is to ask her to justify their inclusion in the forecast and to assess to what extent this justification can be considered even partially correct (e.g., for one of the factors the outcome is a reasonable one, but this is not true not for the combination of the two factors). An instruction-based assessment system should include a comprehensive set of assessment strategies, focusing on different knowledge levels, such as the case illustrated here.

(h) The assessment results are being communicated to the student as part of the dialog. That would have been a merely technical feature if not viewed within the framework of the principles guiding the development of the system. The main objective in returning relevant information to the student is the creation of feedback loops where she is in a position to review her answers, to change the forecast and even to review the original weather data on which the forecast was based. When that happens, the assessment procedures are triggered again and a new cycle takes place.

This section discussed the main prerequisite features of an instruction-based assessment system from the user's perspective (or from the result-on-the-screen perspective). In the next sections we will organize and present these features in a systematic way, referring first to the design rationale of the system, then to its modules, and finally to its implementation (both at the student and teacher levels) and research agenda.

The Design of Instruction-based Assessment in the Weather Lab

The Weather Lab is an instruction-based assessment tool that offers a rich environment for the exploration and simulation of weather phenomena at different stages of the instructional sequence. For example, it can be used after the student learns about each observable weather component (e.g., temperature, wind) and its measurement, to explore how variations in a given component affects weather. It can also be used to simulate multiple-variables combinations after the student learns about unusual weather phenomena Additionally, the student can test her hypotheses about given weather situations by running different combinations of weather components. In the coming sections we present a brief overview of the model underlying the design of the system.

Instruction loops

The working session is conceived as a collection of Instructional Blocks (IB), as shown in Figure 6. Assessment information is generated through the student's interaction with the system. At the end of each block this information is communicated to the student (e.g., in excerpt 3, the system communicates to the student the result of the analysis of the causal relationships between a weather component and a forecast component as stated by the student) and stored within the system (this feature will be clarified in the section that describes the system's structure). This information may eventually affect either the way the student chooses to continue her work (in the example, it may lead to the revision of the questioned forecast component) or the activation of a new instructional block by the system.

Insert Figure 6 about here

Each working session has a focus, or Instructional Goal, which is defined by the teacher or the student. One goal could be to assess the student's ability to understand how variations in given weather components affect weather phenomena. Another goal could be to evaluate the student's ability to infer what configuration of weather components can be correlated with an extreme weather phenomenon.

According to the defined instructional goal, an opening instructional block is activated. Once completed, this instructional block could be followed by a number of new blocks. For example, following the system's assessment of her forecast, the student may choose to consult the Weather Lab about aspects of it questioned by the system. This will start a new instructional block in which the student formulates questions and the system generates answers. Once again, new assessment information is generated leading to a new loop. The session proceeds following this cyclical path until its end.

At the end of the session a summary of the assessment information is generated and communicated to both the student and the teacher, and the system's internal model of the student is updated. This summary information is expected to affect both the student's and teacher's decisions about the next stages in the instruction (The summary report is described in a later section.).

Knowledge levels.

The assessment focuses on three different levels of the student's understanding and performance. The first is the factual knowledge level, including declarative knowledge such as weather facts, weather measurement scales and units, and forecast components. The second knowledge level refers to knowledge of causal relationships between individual weather components and weather phenomena. The third knowledge level focuses on a systemic view of weather phenomena and its causes. An instructional block at this level could be to present an extreme weather phenomenon, such as a tornado, and ask the student to analyze it and infer what different sets or combinations of weather components could lead to its occurrence.

Assessment strategies.

A third element in our design rationale is the repertoire of assessment strategies included in the system. The following are some examples of strategies that differ in nature and in the level of knowledge they address.

The first assessment strategy consists of the 'plain assessment' of the students inputs. For example in excerpt 1, the system assesses the weather data the student had initially entered and, as a result, communicates that there is an irregularity with the temperature mark. This type of assessment is of the form of an input-evaluation-output procedure. Its most plausible implementation is for assessing the student's understanding and manipulation of factual knowledge.

A second example of an assessment strategy is the 'ask-for-factors' strategy. In this case the system first asks the student to indicate the factors which had led her to a given conclusion. Next it runs its own inference mechanisms on the inputted factors and assesses the plausibility of drawing from these factors conclusions like those indicated by the student. Finally, the results of the assessment are communicated to the student. (An example of this strategy appears in excerpt 3.)

A third type of assessment strategy, different in complexity, is the 'question-assessment' strategy, aimed at assessing the student's knowledge state by analyzing her questions to the system. It is obvious that there is no reason to activate this strategy every time the student consults the system. The relevant context for its implementation is when the student asks a question while performing another task, using the system's expertise as an aid or help for accomplishing the task. In this case the system will evaluate the level of the question (e.g., factual or conceptual) and its content, and will assess its relevance to the task. This information will be summarized and included in the report delivered to the student and the teacher at the end of the session.

These assessment strategies are implemented in the actual version of the system. As our research project develops, additional strategies will be added to the system's repertoire.

The Instruction-based Assessment Model

The confluence of the above mentioned elements (namely, instruction loops, knowledge levels, and assessment strategies) into the underlying model of the system can be represented as shown in Figure 7. Each working session has a focus or instructional goal. The session proceeds as a series of instructional blocks. Each new block is implemented on the basis of student or system decisions, following the assessment information generated within the previous block. The format of the activity for each block is chosen by the student or the system from the set of allowed activities or functions (e.g., to consult the system or to initiate a dialog section). The nature of each new instructional block is determined by the assessment strategy being implemented and the knowledge level addressed. Through the working session the assessment information is communicated to the student (as part of the interaction dialogues) and stored within the system. At the end of the session the assessment summary is communicated to the student and the teacher, and stored in the system.

Insert Figure 7 about here

The Weather Lab as a knowledge-based system

Knowledge is a central component in the Weather Lab. Knowledge about weather facts, about weather forecast generation, about the character of the student's responses, and about what kind of dialog to present next are examples of the different kinds of knowledge that enable the system to function as an instruction-based assessment tool. Thus the Weather Lab is a knowledge based system (Clancey, 1987; Davis & Lennat, 1982) being developed following the design principles and methodology which characterize a broad family of educational tools used for tutoring, training, coaching, and as decision-support systems (Polson & Richardson, 1988; Wenger, 1987). A detailed description of the system's architecture is beyond the scope of this paper, but we will briefly refer in the coming sections to its different knowledge modules.

Weather knowledge

The system's weather knowledge-base focuses on the description and prediction of weather phenomena on the basis of observable and measurable surface weather components, like temperature, wind speed and direction, sky conditions, and barometric pressure (Ahrens, 1988). This weather expertise module is a compound of both declarative and procedural knowledge.

The declarative knowledge is comprised of two sets of data:

(a.) Weather facts: This is a collection of numerical and factual data of different types, such as precipitation tables for different geographic regions through the year and temperature maximal and minimal from all over the world. An example of the use of this data appears in excerpt 1, where information about high temperatures recorded in different regions is included in the system's output following the assessment of the data set by the student.

(b) Forecast association lists: This is the main knowledge used by the system for generating the forecast or analyzing the student's forecast. Each list consists of the whole set of possible weather outcomes for a given weather component, in the form:

<weather component> »»» outcome 1, outcome 2, ..., outcome n

For example, for the weather datum "high pressure falling", the following is the set of possible weather outcomes: Fair, rising temperatures, increasing cloudiness, increasing winds, possible showers within 12 hours, rain or snow within 24 hours. Note that this is not the actual forecast, but the set of weather phenomena associated with given pressure conditions. The system contains about 30 forecast association lists, regarding varied wind, pressure, sky, and temperature conditions.

The system's procedural knowledge allows it to generate forecasts based on the input data entered by the student, as exemplified in excerpt 2. A first version, or row forecast, is generated by intersecting relevant sets (association lists) of potential weather outcomes for each item of the weather input data (e.g., wind, pressure, temperature). But this row version may contain redundancies, inconsistencies or contradictions among components (e.g., "precipitation ending" and "possible showers within 12 hours") or among them and some of the input weather data (e.g., "snow within 12 hours" during summer months). A set of procedures is activated to remove these inconsistencies or contradictions (e.g., if "heavy rain" and "heavy snow" are given as two possible outcomes, the contextual temperature data will guide the system's selection of one of these).

The declarative knowledge in the system is represented in a frame-network structure (Anderson, 1988; Collins, Warnock, Aiello & Miller, 1975; Minsky, 1975), while the procedural knowledge is mainly organized as a rule base (Davis, Buchanan & Shortlife, 1977; Winograd, 1982; Winston, 1984).

Assessment knowledge

The Weather Lab's assessment module evaluates the student's inputs and maintains an internal representation of her knowledge state at different stages during the instructional dialog. The two main components of this module are the student model and the assessment mechanism (VanLehn, 1988).

The Student Model. Two kinds of representations of the student's performance are maintained. The first is a working model of the student's state of knowledge in the form of a series of parameters that are constantly revisited and updated during the instructional session. These include: the quality of the data entered by the student, differences between hers and the computer generated forecasts, and the kinds of questions asked by her. The second representation of the student's performance is stored in the form of a session protocol, including all of the student's inputs, the student's activation of interface features (e.g., "clicking" to select a given option), the computer outputs, and key commentaries where needed (e.g., "the student opens the 'forecast formulation' dialog area"). The excerpts considered in a previous section were extracted from such a recorded protocol.

The assessment procedures. These are activated as needed through the instructional dialog. For example, the weather data entered by the student is evaluated at the beginning of the session (as in excerpt 1) looking for abnormalities (e.g., -30˚ F in July in a southeastern state), or inconsistencies. Other procedures will assess the forecast formulated by the student (e.g., her considerations about causal relationships between the data and a given forecast component), as shown in excerpt 3.

Session organization knowledge

This function controls the flow of the session, implementing its various stages according to the plan previously fixed or according to the nature of actual student interactions. In the sample dialog shown in the excerpts, the session resulted as shown in Figure 8.

Insert Figure 8 about here

The sequence reflects the default session plan built-in in the system. But a more likely situation would be that the session plan would be determined by the teacher or the student, focusing on a specific issue or instructional goal. For example, if the goal is to explore how modifications of the weather components affect an unusual or extreme weather situation (e.g., a hurricane), the student activity will focus mainly on entering and modifying input data, running the computer forecasting procedures to simulate the unusual weather, and consulting the computer knowledge base (query mode).

The session plan (either determined by the teacher, the student, or by default) could be altered in many ways. One way it might be altered is as a result of the student's decisions in response to the system's assessment outputs. For example, in excerpt 1, after assessing the temperature entered by the student as unlikely, the system suggests the option of revising this data item. If the student decides to do that, the current dialog path is interrupted; its next configuration will depend on the nature of the modifications done by the student to the input data.

The default session plan could also be modified when a particular dialog situation requires a change in assessment strategy. Suppose that the comparison stage shows that there are no differences between the student's and the computer generated forecasts. In that case the strategy focusing on asking about forecast items not included in the computer's forecast is no longer relevant, and will be replaced by a next-level strategy for assessing the student's handling of causal relationships. For example, the system may ask the student to consider what changes in the formulation of the forecast will occur following a change in the value of a given input item (e.g., pressure level).

Assessment Report Generation

Informing the student's and teacher's decisions is a main objective guiding the design of the system. Besides delivering continuous feedback throughout the instructional session, the system generates a summary report at the end of the session. The nature of the report to be delivered to the student and the teacher is currently being defined. It will be based on the information contained in the student model, and depending on the instructional goal of the session, it may focus on one or more of the following aspects of the student's performance (see Figure 9). First, it may focus on her knowledge of factual information, both about the weather components and weather phenomena. Second, the report may indicate her ability to identify causal relationships while explaining or predicting weather phenomena (e.g., to formulate a forecast based on observed data, or to analyze the possible causes of a snow storm). Third, it may report her questioning strategies, based on the level of the questions (e.g., facts, causes, weather patterns) and the context for their formulation (e.g., while entering data, formulating the forecast, or justifying an item in her forecast). Finally, recommendations about aspects of the student's performance that should be addressed in future sessions will be included in the report.

The entire protocol recorded throughout the session is also avaiLable for printout, whenever the teacher considers it necessary.

Insert Figure 9 about here

Implementation and Research agenda

A pilot version of the Weather Lab is now ready to be implemented. In this final section we will describe different aspects of our implementation plan.

The instructional framework will be completed by developing a sequence of learning activities. Each activity will focus on what we called the instructional goal for a session. For example, understanding how the combined effects of two observable variables such as wind direction and sky conditions influence weather could be an instructional goal for a session. This instructional sequence will proceed gradually from learning about individual weather components, through manipulating and combining increasing number of weather variables, to hypothesizing about, and simulating, a multi-variable weather phenomenon.

Our research agenda pursues three main objectives. The first is the improvement of the system. In particular we are interested in refining the student modeling procedures and in expanding the assessment strategies repertoire to cover as many different aspects of the student's knowledge state as possible.

The second objective is to learn about the cognitive processes triggered by interactions with the system. The student is involved in cognitively demanding tasks at different stages, such as to indicate causal relationships between weather components and outcomes, to formulate a hypothesis about the possible outcome from a particular configuration of multiple weather components, or to consult the system to obtain information that is relevant to the task being performed. We are interested in what types of conceptual models of weather processes the student has, and how these models are affected (e.g., completed, improved, or confused) and evolve through her work with the Weather Lab.

As part of our interest in cognitive processes, particular attention will be given to an analysis of the questions the student asks to the system, their content and context for their formulation. We are interested in identifying questioning patterns, in assessing how these patterns reflect the student's knowledge state, and in tracing if (and how) these patterns change as the work with the system progresses.

A third focus in our research agenda relates to the ways assessment information is used by the student and the teacher to inform instructional decision making. We are interested in identifying patterns of actions and their causes at the micro-level within a session (mainly as a result of the student's decisions), at the 'between-the-sessions' level (both by the student and the teacher), and at the macro-level, affecting the next cycle of the instruction (mainly as a result of the teacher's decisions). Part of our future work at this level will also focus on evaluating and improving the ways the assessment information is communicated (e.g., immediate feedback, session report), both in its format and content.

At the end of the research and implementation process we expect to have a better understanding of the process of developing an instruction-based assessment tool, of the cognitive implications of the work within such a learning environment, and of the way assessment information is used by the student and the teacher to make decisions about the instruction. As a result, we hope this knowledge will contribute to generalize the presented approach to the development of instruction-based assessment tools in other content areas.

References

Ager, T. (1990). From interactive instruction to interactive testing. In R. Freedle (Ed.), Artificial intelligence and the future of testing. Hillsdale, NJ: Erlbaum .

Ahrens, D. C. (1988). Meteorology today. St. Paul, MN: West Publishing Company.

Anderson, J. R. (1988). The expert module. In M. Polson & J. J. Richardson (Eds.), Foundations of intelligent tutoring systems. Hillsdale, NJ: Erlbaum.

Bachor, D. G. (1990). Toward improving assessment of students with special needs: Expanding the data-base to include classroom performance. Alberta Journal of Educational Research, 36(1), 65-77.

Brown, A. L., Ash, D., Rutherford, M., Nakagawa, K., Gordon, A., & Campione, J. (1993). Distributed expertise in the classroom. In G. Salomon (Ed.), Distributed cognitions, (pp. 188-228). New York, NY: Cambridge University Press.

Burton, R. B. (1982). DEBUGGY: Diagnosis of errors in basic mathematical skills. In D. Sleeman & J. S. Brown (Eds.), Intelligent tutoring systems. London: Academic Press.

Burton, R. B. (1988). The environment module of intelligent tutoring systems. In M. C. Polson & J. J. Richardson (Eds.), Foundations of intelligent tutoring systems. Hillsdale, NJ: Erlbaum.

Carbonell, J. C. (1970). AI in CAI: An artificial intelligence approach to computer-assisted instruction. IEEE Transactions on Man-machine Systems, 11(4), 190-202.

Clancey, W. (1987). Knowledge-based tutoring. Cambridge, MA: MIT Press.

Cognition and Technology Group at Vanderbilt (1996). Multimedia environments for enhancing learning in mathematics. In S. Vosniadu, E. de Corte, R. Glaser, & H. Mandl (Eds.), International perspectives on the design of technology supported learning environments (pp. 285-305). Hillsdale, NJ: Erlbaum.

Collins, A., Hawkins, J. & Frederiksen, J. R. (1993). Three different views of students: The role of technology in assessing student performance. The Journal of the Learning Sciences, 3(2), 205-217.

Collins, A., Warnock, E. H., Aiello, N. & Miller, M. L. (1975). Reasoning from incomplete knowledge. In D. G. Bobrow & A. Collins (Eds.), Representation and understanding. NY: Academic Press.

Davis, R., Buchanan, B. G.,& Shortlife, E. H. (1977). Production rules as a representation for a knowledge-based consultation system. Artificial Intelligence, 8, 15-45.

Davis, R., & Lennat, D. B. (1982). Knowledge-based systems in artificial intelligence. New York: McGraw-Hill.

Deringer, D. K. (1986). Technology advances that may change test design for the future. In The redesign of testing for the 21st Century, proceedings of the 1985 ETS invitational conference. Princeton, NJ: ETS.

Freedle, R. (Ed.). (1990). Artificial intelligence and the future of testing. Hillsdale, NJ: Erlbaum.

Glaser, R. (1986). The integration of instruction and testing. In The redesign of testing for the 21st Century, proceedings of the 1985 ETS invitational conference. Princeton, NJ: ETS.

Goldstein, I. P. (1982). The genetic graph: A representation for the evolution of procedural knowledge. In D. Sleeman & J. S. Brown (Eds.), Intelligent tutoring systems. New York: Academic Press.

Gong, B., Venezky, R. L. & Mioduser, D. (1992). Instructional assessments: Lever for systemic change in science education classrooms. Journal of Science Education and Technology, 1(3), 157-176.

Harris, K. H., & Longstreet, W. S. (1990). Alternative testing and the national agenda for control. Social Studies, 81(4), 148-152.

Lesgold, A., Ivill-Friel, J. & Bonar, J. (1989). Toward Intelligent Systems for Testing. In L. B. Resnick (Eds.), Essays In honor of Robert Glasser. Hillsdale, NJ: Erlbaum.

Lipson, J. (1988). Cognitive and environmental perspectives on assessing achievement. In Assessment in the service of learning : proceedings of the 1987 ETS Invitational Conference. Princeton, NJ: ETS.

Lock, R. (1990). Assessment of practical skills. Part 2: Context dependency and construct validity. Research In Science and Technological Education, 8(1), 35-52.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Eds.), The psychology of computer vision. NY: McGraw-Hill.

Polson, M. C., & Richardson, J. J. (1988). Foundations of intelligent tutoring systems. Hillsdale, NJ: Erlbaum.

Snow, R. E. (1988). Assessment in the service of learning. In Proceedings of the 1987 ETS invitational conference. Princeton, NJ: ETS.

Treffinger, D., Isaksen, S., & Dorval, K. (1994). Creative problem solving: An overview. In M. Runko (Ed.), Problem finding, problem solving, and creativity, (pp. 223-236). Norwood, NJ: Ablex.

VanLehn, K. (1988). Student modeling. In M. Polson & J. J. Richardson (Eds.), Foundations of intelligent tutoring systems. Hillsdale, NJ: Erlbaum.

Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., Mislevy, R. J., Steinberg, L. & Thissen, D. (1990). Computerized adaptive testing: A primer. Hillsdale, NJ: Erlbaum.

Wang, M. (1988). The wedding of instruction and assessment in the classroom. In Proceedings of the 1987 ETS invitational conference. Princeton, NJ: ETS.

Ward, W. (1986). Measurement research that will change test design for the future. In The redesign of testing for the 21st century: Proceedings of the 1985 ETS invitational conference. Princeton, NJ: ETS.

Wenger, E. (1987). Artificial intelligence and tutoring systems: Computational approaches to the communication of knowledge. Los Altos, CA: Morgan Kauffman.

Winograd, T. (1982). What does it mean to understand language. In D. Norman (Eds.), Perspectives on cognitive science. Hillsdale, NJ: Erlbaum.

Winston, P. H. (1984). Artificial intelligence. Reading, MA: Addison-Wesley.

Figure Captions

Figure 1 : The Weather-Data Input Environment

Figure 2 : The Dialog Environment

Figure 3: Excerpt 1: Assessing the initial weather data

Figure 4: Excerpt 2: The system's subject matter expertise

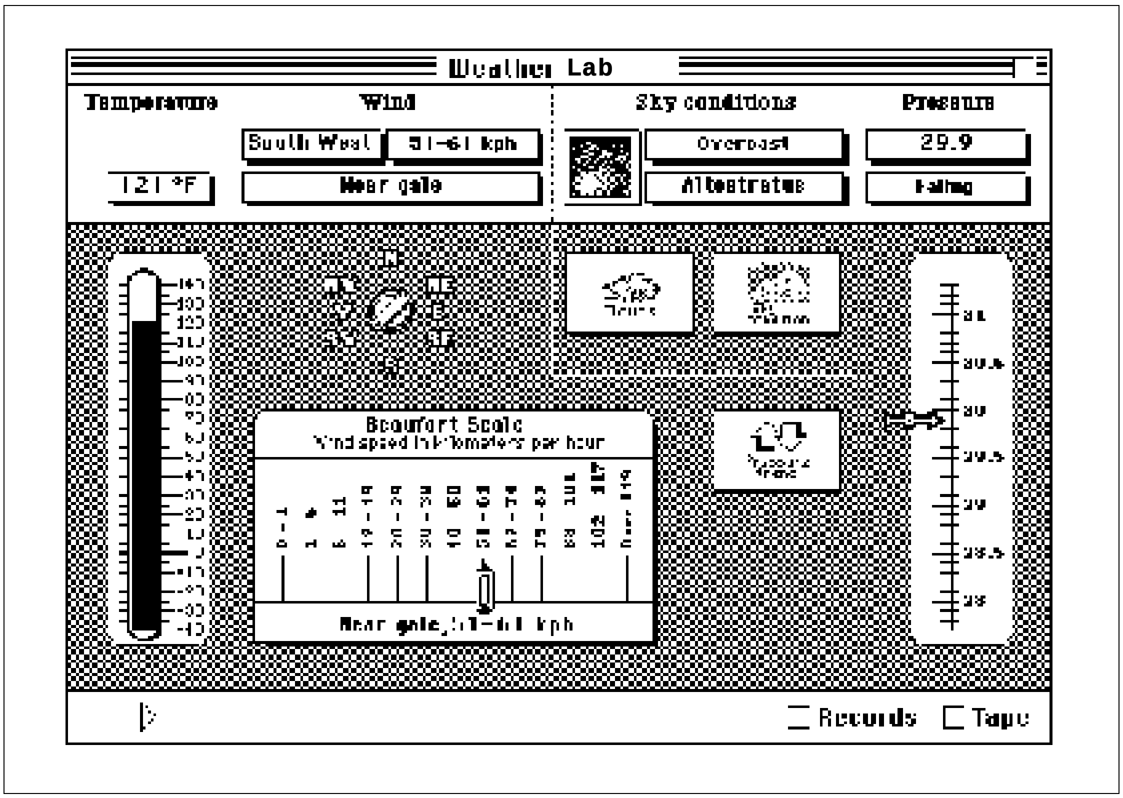
Figure 5: Excerpt 3: Assessment of the understanding and application of causal relationships

Figure 6: Structure of a session

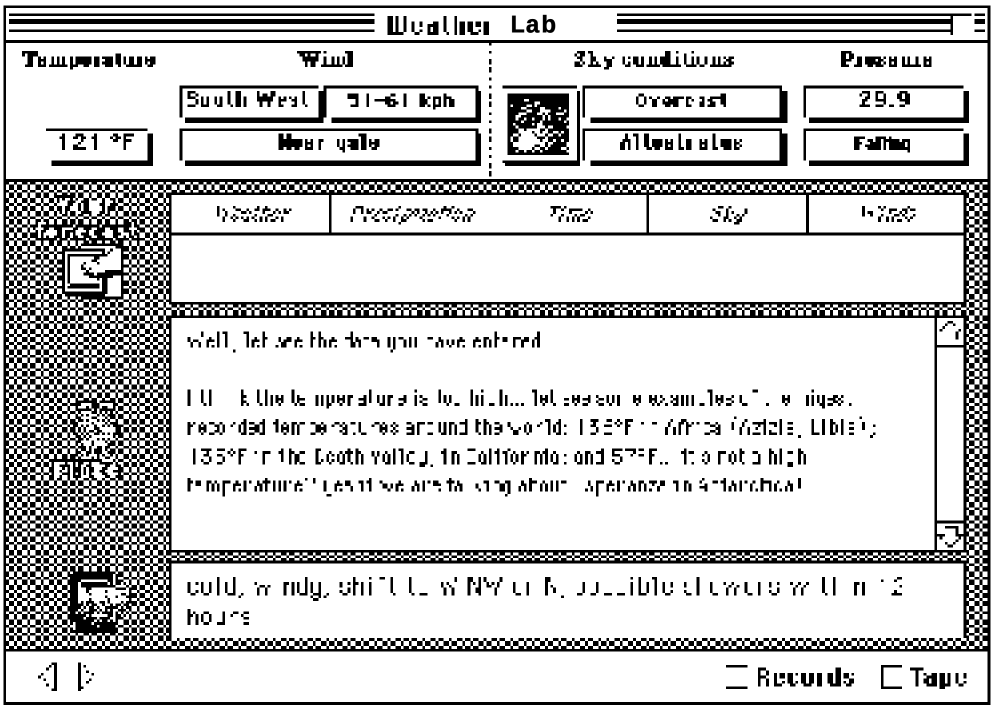
Figure 7: The Instruction-based Assessment model

Figure 8: Sequence of Instructional Blocks appearing in Excerpts 1 to 3

Figure 9 : Aspects of the Student's Performance which may be Included in the session report



Weather Lab
Temperature
Wind
Sky conditions
Pressure
South West
51-61 kph
Overcast
29.9
121 °F
Near gale
Altostratus
Falling
Beaufort Scale
Wind speed in kilometers per hour
Near gale, 51-61 kph
Records
Tape


Weather Lab
Temperature
Wind
Sky conditions
Pressure
South West
51–61 kph
121 °F
Near gale
Overcast
Altostratus
29.9
Falling
Weather
Precipitation
Time
Sky
Wind
Records
Tape

c: Well, the data are now:

May
Wind: SW
Pressure: 29.9, Falling
Sky conditions: Overcast, Altostratus
Temperature: 128

c: Let's check the data you have entered...

I think the temperature is too high... let's see some examples of the highest recorded temperatures around the world: 136°F in Africa (Azizia, Libya); 135°F in the Death Valley, in California; and 57°F... that's not a high temperature except for Esperanza in Antarctica!

Now, do you want to revise the temperature data?
If that is the case please simply modify your data in the input window.
If not, be aware that this information (128˚F) is part of the data we are going to use for further work.

c: Please enter your forecast
(click on button 'Your forecast')
thank you !

(student clicks on 'Your forecast' - opening the forecast formulation dialog area)
(student clicks on 'Weather' - enters a general weather component)
(student clicks on 'Sky condition' - enters a specific component)
(student clicks on 'Your forecast' - closes the forecast formulation dialog area)

c: Good, your forecast is now:

cold, clearing, little change

c: Well, before we continue I have to generate the computer forecast. Please wait a few seconds. The forecast will appear below.

The computer's forecast is now:

cold, windy, shifting to W NW or N, possible showers within 12 hours

c: You mentioned in your forecast "clearing";
that possibility does not appear in the computer's forecast.
On which of the following weather components is your forecast based?
(Please select one or more components)

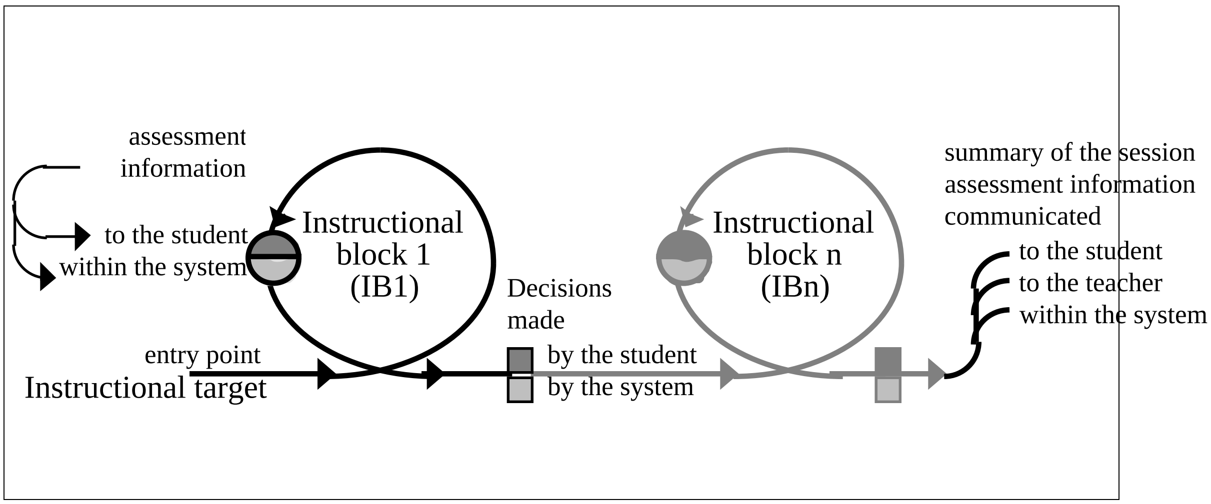
s: Wind, Sky conditions

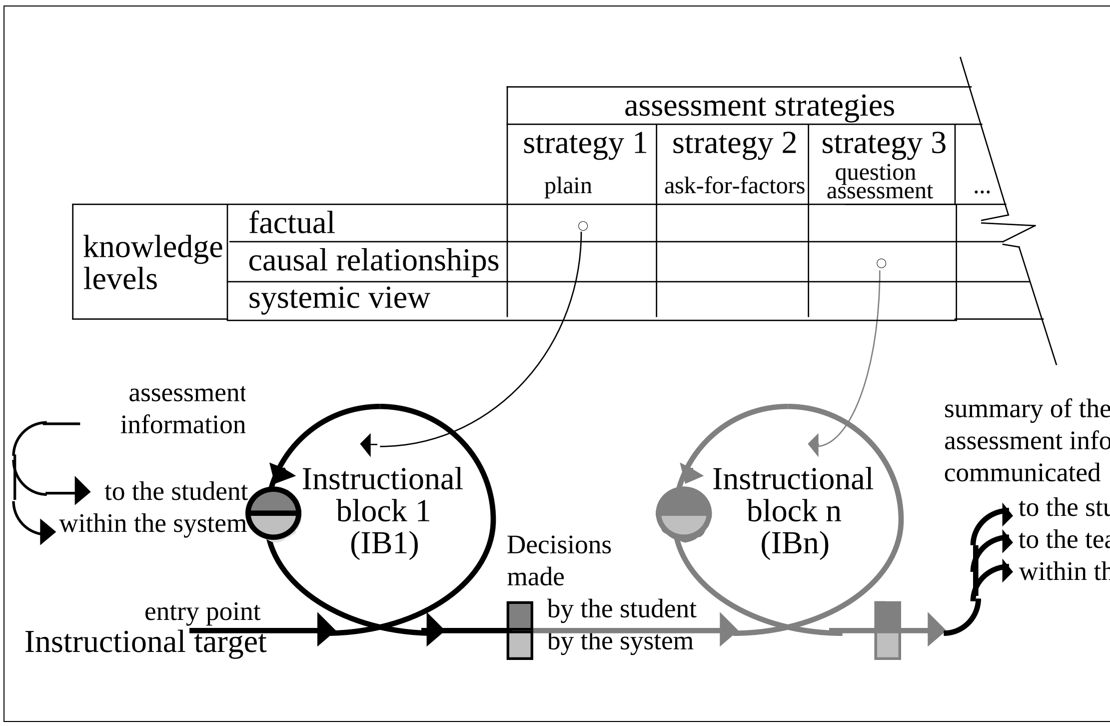
c: O.K.,
For Wind: SW
"clearing" could be a possible outcome
For Sky conditions: overcast
"clearing" is not a reasonable outcome
For Wind & Sky conditions combined
"clearing" is not a reasonable outcome

c: Now, you also mentioned in your forecast " little change";
that possibility does not appear in the computer's forecast.
On which of the following weather components ...
... *(section omitted)*

c: The following weather components appear in the computer's forecast but not in yours:

windy, shift to W NW or N, possible showers within 12 hours


assessment
information
to the student
within the system
Instructional
block 1
(IB1)
entry point
Instructional target
Decisions
made
by the student
by the system
Instructional
block n
(IBn)
summary of the session
assessment information
communicated
to the student
to the teacher
within the system


assessment strategies
strategy 1
plain
strategy 2
ask-for-factors
strategy 3
question assessment
...
knowledge levels
factual
causal relationships
systemic view
assessment information
to the student
within the system
Instructional block 1 (IB1)
Instructional block n (IBn)
Decisions made
by the student
by the system
entry point
Instructional target
summary of the session assessment information communicated
to the student
to the teacher
within the system

	Focus	Detail
Stage 1	**Assessment of input data**	• assessment of unlike data • assessment of inconsistent data
Stage 2	**Forecast generation**	• student forecast input • computer forecast generation
Stage 3	**Forecast comparative evaluation**	• forecast components included by the student but not by the computer • forecast components included by the computer but not by the student
Stage 4	**Causal relationships between weather and forecast components in the student's forecast**	• input of the factors considered by the student to include a forecast component • student's factors assessment • feedback generation

Factual Knowledge		Weather components
		Weather phenomena
Causal Relationships	From Components to Phenomena	By variable
		Variables combined
	From Phenomena to Components	By variable
		Variables combined
Questioning ability		Context
		Level