

Towards an Ontology for Alternative Assessment in Education*

Steven L. Tanimoto

Box 352350

Dept. of Computer Science and Engineering
University of Washington, Seattle, WA 98195

September 25, 1998

(C) Copyright 1998, Steven L. Tanimoto

Abstract

A set of guiding principles is stated for the development of standards for representing student educational assessment information. Unlike traditional academic transcripts, the items in portfolios, subjective written evaluations, self-assessments, and computer-based records of activity are complex representations of student achievement, involvement, or inclinations. Even more so than traditional grades, they depend upon a great deal of contextual information in order to be interpreted in useful and reliable ways. This paper also identifies the essential informational components of alternative assessment records and suggests a standard form for their representation. The consideration of evidence, judgment, context, and justification, as described in this paper, is relevant to the improvement of conventional (e.g., multiple-choice test) assessment methodologies, as well.

*First presented at the June 1998 meeting of the IEEE Learning Technology Standards Committee held in Pittsburgh, PA. Please send comments and corrections to Steve Tanimoto, Dept of Computer Science and Engineering, Box 352350, University of Washington, Seattle, WA 98195; email: tanimoto@cs.washington.edu

1 Introduction

The efforts of several groups to enable interoperability among electronic learning environments include those of the IEEE Learning Technology Standards Committee and Educom's Instructional Management System [EDUCOM 1998]. These efforts are motivated in part by the opportunity for online learning presented by the Internet and the World Wide Web, and by various shortcomings of existing software. With the advent of electronic learning environments, including intelligent tutoring via computers and computer based construction environments such as Interactive Physics, Rocky's Boots, and many others, part of a student's getting started with a new piece of software involves the program learning about the student's current state of knowledge. This information allows the program to present material or suggest activities that are appropriate for the student. The time spent acquiring this information could be minimized if it were all available in one or more data files constructed during the student's earlier interactions with other educational programs. In addition, the existence of a thorough, machine-readable record of the student's education creates the possibility of having many separate (and possibly diverse) software programs that contribute to the student's education in a coordinated way.

The different programs may handle different aspects of the educational process. Some programs may be oriented towards the presentation of new material. Others may focus on assessment of the student's comprehension of certain material, while yet other programs may perform customized curriculum planning for the student. The complete record of a student's learning (let's call it the student's dossier) could be readily processed to produce resumes, career planning analyses, or progress reports for specific subjects or time periods. The CAI programs that utilize and add to the student dossier may utilize not only present and future computer technology, including future multimedia and virtual reality technologies, but also future pedagogical methodologies.

Not only does the student dossier represent the educational experiences of the student obtained during sessions with CAI programs, but it may also contain assessments of non-computer-based learning experiences, obtained via experts (e.g., teachers) and interview programs. (See [Tanimoto 1992] for additional justification for such dossiers.)

Current work on learner models captures many of the key ideas regarding the contents and purposes for dossiers (e.g., see [Murphy and McTear 1997]). However, little has yet been said about how non-test-based assessment items should be handled. (For a justification of non-test-based assessment see [Hoffman 1962], and for a description of conventional and alternative assessment methods see [Linn 1989] and [Broadfoot 1986], respectively.) The purpose of this paper is to give the broad outline of an approach for handling of alternative assessment information within student learning databases.

Part of the philosophy behind this paper is that an evaluation of student

learning should always be qualified in terms of the particular evidence for that evaluation, whether the evaluation is done by a human teacher, a computer testing program, or through a self- or peer-evaluation process. This approach to representing the student's education allows important decisions to be made where any possibilities for error can be appropriately taken into account, and additional evidence gathered when appropriate.

The primary purpose for the assessment information of concern in this paper is to improve the efficiency of the student learning process. This includes helping solve the "transfer problem" which occurs whenever a student moves from one educational environment to another; most of the information built up about the student in the first environment is either lost or simply not transferred to the second environment, so that time and energy must be wasted in rediscovering the precise needs of the student.

2 Principles

The following list of six principles is offered to guide the development of standards for electronic records of assessment-related information in student records.

1. Multiple forms of assessment (e.g., multiple-choice tests, evaluations of students' prose, log-file analyses) are essential in developing an accurate model of a students' knowledge, skills, learning styles, experience, achievement, motivation, and goals.
2. Any single assessment item (e.g., test result, project evaluation, etc.) has a degree of unreliability and a degree of incompleteness that bears on its usefulness in analysis.
3. The effects of uncertainty and incompleteness in an analysis of the student can be reduced by recording and considering information about the context in which the assessment was performed.
4. When possible, assessment information should be evaluated in a manner similar to that used when evaluating evidence, using methodologies appropriate to handling of evidence, including using means to determine the reliability of evidence, and using means to make logically or mathematically valid inferences from the evidence.
5. Assessment information, like many other kinds of information, may be owned by people or institutions. However, the parties to an assessment event have various rights to the information about the event. A system for representing assessment information must contain provisions both for the protection of ownership and for the exercise of these rights.

6. The cost of storage media, per unit of information, continues to fall. To the extent practicable, summaries or evaluations of student performance should be accompanied by detailed digital records of the student's actual activity. This may permit explanation, verification, and/or reworking of the assessments, if the objectives or methodologies for the assessments should change in the future.

3 Taxonomy of Assessment Items

The information pertaining to alternative assessment can be classified into the following three general categories: (1) educational history items, (2) portfolio items, and (3) evaluation items. The history items represent or reflect some actual activity that the student was involved with. The portfolio items are representations of constructions and accomplishments by the student that can be presented to other people as samples of the student's work or as parts of an explanation of the student's educational background. Although intended for viewing by people, portfolio items also contain machine-interpretable descriptions. Evaluation items include the results of inferences and judgments about the student's understanding, motivation, skill levels, etc. These items generally identify evidence, method of inference, and conclusions, including indications of doubt and degree of reliability.

3.1 Educational History Items

1. Registration record (statement that student registered for a particular activity at a particular time and date)
2. Activity Log (sequence of events)
3. Activity description (the learning activity and software)
4. Event description (meanings of events)
5. Communication log (email, newsgroup postings, audio calls, etc.)
6. Test, quiz, or exam and student answers

3.2 Portfolio Items

1. Copy of project report
2. Pointer, URL or reference to actual project
3. Journal or Notebook or pointer to such

3.3 Evaluation Items

1. Evaluation of project or activity
2. Evaluation of understanding or skill
3. Evaluation of motivation or learning styles
4. Analysis of student performance on a test
5. Evaluation of educational progress
6. Result of an advising session

This taxonomy is organized according to degrees of summarization in the following sense. The details of participation in an activity are in a sense condensed and summarized by the final student product of the activity. For example, a student's participation in a chemistry experiment in an online simulated laboratory is represented by a log of the student's actions in each phase of the experiment. However, this activity is summarized in the student's lab report, which is a document that might form part of the portfolio. Finally, this document, after analysis by a teacher or agent, is represented in an even more condensed form in an evaluation.

In this case, and ideally in all cases, for each activity in the student's educational history, there is one or more corresponding portfolio item, and in turn, corresponding to that or those, evaluation items. Thus there are relationships that connect assessment records across the top-level classes of the taxonomy.

Another distinction that one can see in this taxonomy is that the educational history items and portfolio items constitute "raw material" for the evaluation items. The portfolio items can be thought of as student products, while the educational history items are typically by-products of the student activity. The evaluations will commonly be products of teachers and programs, but in the case of self-evaluation, one would have an evaluation item that also happens to be a product of the student.

Of course, various categories of activity could be identified and the taxonomy thus refined. However, the intention of this paper is to provide only the general outline of this structure.

4 Components of Assessment Records and Files

Now that a general taxonomy has been established, let us proceed with some more concrete forms of representation.

A representation method is described having two levels: the file and the record. An assessment file is intended to represent the results of any specific assessment process such as computer analysis of an essay by the student or one or more

CAI sessions with the student. An assessment file contains a context header, a collection of assessment records, and a list of “post requests” to a summary file. The context header provides information that applies or that may apply to all of the records in the file. Each record describes either a history item, a portfolio item, or an evaluation item. The assessment file may contain references to other files or information resources. These may be local (within the same educational profile) or they may be external (for example in the database of an academic institution at a remote web site).

The list of post requests is a representation of summary data that derives from the other information in this assessment file and which should be used to update an overall summary file for the student. While it may seem redundant to include this information here rather than to simply transmit it to the summary record, there are two advantages to storing it here: (a) the influence of the activities represented in this file upon the overall summary record is clear and can be interrogated, and (b) it is possible to conveniently create this file while disconnected from the central summary file and yet have the influence communicated at such time as the connection may be made.

4.1 Context Header

An assessment file begins with a context header. This contains information that applies or potentially applies to all of the records in the file.

1. Date and time created
2. Id of root file (reference to core file of student’s dossier that contains student id., info access policies, etc.)
3. Name and version of application program (e.g., intelligent tutor program) writing this record
4. Name and version code of assessment standard to which this file conforms.
5. Name(s) and version(s) of domain model(s) referenced, if any.
6. Date and time of last update to this file.
7. Number of assessment items in this file.
8. Special conditions known to be in force during creation/updating of this file. E.g., student not working alone, student using various tools, noisy or nonideal environment, etc. Information here applies to all records in file unless explicitly overridden in individual records.

4.2 Assessment Records

Within an assessment file are zero or more assessment records. Each record describes a history item (participation by the student in an activity), a portfolio item (a product of a learning activity), or an evaluation (set of judgments about the student's learning).

The number of these records is given in the context header above.

Each assessment record contains some or all of the following components:

1. type of record (e.g., activity log, activity description, portfolio entry, etc.)
2. reference to file, URL, or external database entry that contains the actual log file, portfolio document, etc.
3. a domain-based concept reference. e.g.,
domain1(mathematics).algebra.linearfn.intercept.computing (applicable to evaluations of understanding and skills).
4. a learning category (exposed to concept, has basic understanding of concept, is able to use in problem solving, is skilled in use, etc.).
5. possible misconceptions and attitude problems (e.g., has misconception: intercept = slope, has aversion or phobia).
6. category of evidence (answered multiple-choice question correctly, incorrectly; gave clear oral explanation, etc.)
7. detail of evidence (e.g., including time spent, testimonial, etc., the text of the multiple-choice question, reference to recorded speech of oral answer, etc.)
8. inferred skill level, level of understanding, facet of concept, etc.
9. assessor's degree of confidence that inference is valid.
10. any special assumptions required for this inference.
11. representation of the reasoning behind this judgment, if available.
12. information about circumstances of this judgment (e.g., limitations on the assessor's time or knowledge, references consulted during the assessment, etc.).
13. assessor's identity.
14. date and time of creation of this record.

4.3 List of Post Requests

When assessment is performed, the results should be recorded in two ways. First, either an assessment record is created and added to an existing assessment file, or a new assessment file is produced and one or more assessment records inserted. Then, summary information from the new or updated file should be carried over to a summary file for the student. The summary file may or may not be directly accessible to the system handling this assessment. For example, if the summary file is on the student's home computer but this assessment is being performed on a university computer, the updates may have to be downloaded at a later time. The post-requests part of the assessment file contains the updates that represent the influence of this assessment file upon the student's summary file.

Here are a couple of examples of possible post requests.

```
Post 3.2 additional hours spent on math.algebra
```

```
Post to math.summary.algebra.mastery at 5.7 with confidence 3.2  
by NetBasedAlgebra3.2, based on 3.2 hours ending 14 April 1995.
```

If this file is created in a session that is disconnected from the summary file, a means is needed to make the update to the summary file as soon as a connection becomes established (which may be any time after the session). The actual mechanism for this is beyond the scope of this paper. However, we note that if this assessment file is changed after an update to the summary file has been performed, another update will be required to correct the original (and any other previous updates).

5 Open Issues

Two issues that emerge from desiderata above are (1) how to handle the questions relating to the distributed nature of the overall student profile: posting of summary information, keeping it updated, and accessing files on remote servers, and (2) what provisions there should be for representing results of evaluations, justifications for conclusions, and assumptions. For example, languages could be provided which take advantage of descriptive methods from probability, symbolic logic, fuzzy logic, and other inference methods in the representation of evaluations. "On the basis of this quiz, there is a 0.7 probability that the student understands the constancy of ratios of corresponding sides of similar triangles."

Two additional issues worthy of attention are (a) the treatment of groups of students including assessment of group work, and (b) considerations for assessments of dynamically changing student work. In the case of a group project, provisions may be needed for evaluations of the work of the group as a whole, as well as evaluations of the contributions of the individuals. There may be a need

to record evidence and reasoning used in determining how to allocate credit (and assessments of learning learning) to particular group members.

In the case of dynamically changing student work, such as a large art project, two approaches are possible: separate assessments of successive snapshots of the project, or repeated revisions to a single assessment as the project develops. The latter approach may seem appealing in its conservation of storage space; however, it poses difficulties in terms of maintaining consistency between assessment and object assessed, and it disposes of potentially valuable history information.

6 Acknowledgements

Thanks to Adam Carlson and David Madigan of the Univ. of Washington and to Steve Ritter of the Dept. of Psychology, Carnegie-Mellon University, for commenting on previous drafts. The writing of this paper was supported in part by NSF Grant CDA-9616532.

7 References

1. Broadfoot, Patricia (ed). 1986. *Profiles and Records of Achievement*. London: Holt, Rinehart and Winston.
2. EDUCOM. 1998. Instructional Management System home page. <http://www.imsproject.org/>.
3. Hoffman, Banesh. 1962. *The Tyranny of Testing*. NY: Crowell-Collier.
4. Linn, R. L. (ed). 1989. *Educational Measurement*, Third Edition. London: Collier Macmillan Publishers.
5. Murphy, M., and McTear, M. 1997. Learner modelling for intelligent CALL. In Jameson, A; Paris, C.; and Tasso, C. (Eds.) *User Modeling: Proceedings of the Sixth International Conference, UM97*. Vienna, New York: Springer Wien New York. pp. 301-312.
6. Tanimoto, S. L. 1992. Beyond the naivety of grades: Educational record keeping for the twenty-first century. Technical Report 92-07-09, Dept. of Computer Science and Engineering, Univ. of Washington, Seattle WA, USA, July.