Tools for Constructing Targeted Feedback in Online Instruction

William Jerome, Office of Technology for Education, Carnegie Mellon University, USA wjj@andrew.cmu.edu
John Rinderle, Office of Technology for Education, Carnegie Mellon University, USA jar2@andrew.cmu.edu
Diana Marie Bajzek, Office of Technology for Education, Carnegie Mellon University, USA
db33@andrew.cmu.edu

Abstract: This paper describes a new assessment software system which allows assessment authors to tightly integrate content with assessment. The system affords significant flexibility, evaluating assessments according to author specified criteria that targets feedback to the students' particular areas of misconception and keeps instructors informed on class progress.

Well designed technology-enhanced learning environments collect a lot of information from the students' activities as they interact with the instructional materials. Assessments need to be carefully designed and presented if the student learning data collected is to constructively inform further development of student learning activities as well as inform instructors about their individual classes. Authors in our online environment take advantage of the assessment creation process and student progress reporting tools to carefully embed authentic assessments within their instruction materials.

Any piece of feedback received by the student will influence his or her learning in some way. It is therefore important that assessment activities, performance measures, and feedback given to students and instructors closely align with the desired learning objectives and be provided in a timely manner. Students should receive feedback on errors immediately whenever possible, and instructors should be able to get real-time information about student interactions with assessment materials.

Description of New Assessment Software System

In 2007 we released a new assessment software system which provides content authors the ability to tightly integrate content with assessment along with significant flexibility constructing assessments that can target feedback to the students' particular areas of misconception. This system replaced an old set of software that was much more traditional in the types of questions it could ask (multiple choice), information it could provide instructors (very basic bar graphs of answers), and most importantly very limited in the way it evaluated student responses. The new system was developed based on the experience gained from the simpler more limited infrastructure in addition to input from existing content authors and instructors.

In the new system, question prompts, hints, and feedback may contain most forms of instructional material including text, images, tables, lists, audio, video, formulae, animations, simulations, etc... Additionally, declarative content can precede, follow or surround questions. This richness allows assessments to serve as instructional materials. For example, our learning environments offer "Learn by Doing" exercises where new material is presented to students through interactive, hands-on guided practice. Built using our assessment software system, we are able to collect learning data for course designers, provide reports on student performance to instructors, and provide highly targeted feedback to students. Even though such content may be delivered through this assessment software system, from the student perspective the presentation is that of a learning activity and not of a formal quiz.

Course Design that Supports Student Learning
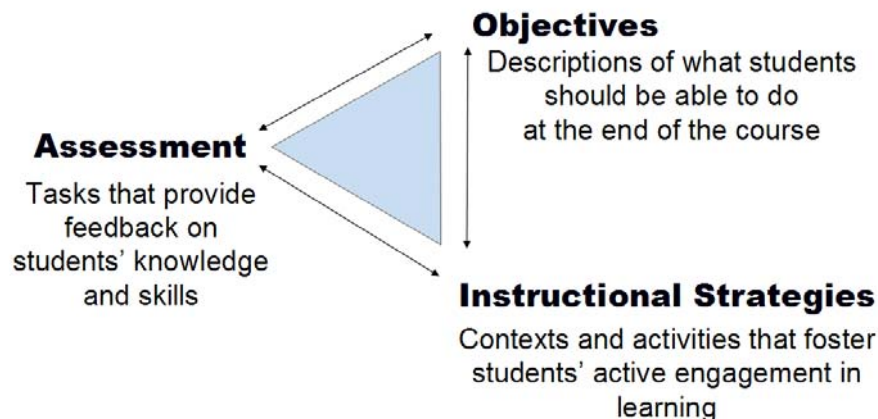
Figure 1: Assessment activities, performance measures collected, and feedback given to students and instructors must closely align with desired the learning objectives.

A question presents the student with one or more inputs or fields where their responses are entered. The way these inputs are displayed depends both on the type of question (using appropriate web-based widgets for the question type) as well as the design of the question (ex: text inputs can follow a prompt or be embedded within it).

The assessment software system natively supports a variety of question types as follows:

Question type	Description
Multiple choice single response multi. responses	A list of choices is given. The student may select a single choice. A list of choices is given. The student may select one or more choices
Ordering	A set of items is given. The student must arrange these items in the proper order
Text	One or many blanks are displayed within the question prompt. The student must supply a word or phrase. Evaluated by string matching or regular expression.
Numeric	One or many blanks are displayed within the prompt. The student must supply a numeric response. Scientific notation and significant figures are supported.
Image Hotspot	An image is displayed with in the prompt. Regions within the image are clickable. The student must select one or more regions that make up their response.
Fill in the Blank	One or many blanks are displayed within the question prompt. The student must select from a list of choices to complete each blank.
Short Answer	A text field is provided where the student may respond with one or two paragraphs in free form. Such questions are generally scored manually by the instructor.
Essay	A large text field is provided where the student may respond with several paragraphs in free form. Such questions are generally scored manually by the instructor.

Figure 2: Question types available in our new assessment software system.

The above list of question types are similar to what one might expect to find in current course management systems. The primary difference comes in how they are evaluated and provide feedback. Reports from the National Science Research Council (NRC, 1996, 2004) have recognized the importance of formative feedback provided by assessment as vital to effective learning environments. We have applied this idea by aligning assessment creation, student feedback and instructor reporting around feedback and categories of student error.

By targeting common misconceptions rather than providing generic comments, course designers can provide feedback better suited to the learning situation. This is particularly valuable for questions which solicit open ended responses, such as questions expecting a numeric or text response. In these situations the student can be provided with information more likely to assist with the current problem solving task at hand. To identify opportunities for feedback, course designers rely upon their teaching experience and domain expertise, examine previous artifacts of instruction, and engage in a process of iterative refinement driven by course data.

Enhanced Evaluation Criteria Provides the Key to Flexible Targeted Feedback

Student responses are flexibly evaluated according to criteria specified by the content author. Each student response is systematically evaluated against sets of author defined criteria which we call response conditions. This process sorts the student response into a group of similar responses. These groups (response conditions) determine the feedback and score the student receives. Content authors construct response conditions that trap common student errors. If the criteria given in the condition match the student response, then the corresponding score and feedback are assigned.

Response conditions also identify to the instructor the frequency of each type of correct or incorrect error. Each response condition may be assigned a label or tag which is used to identify the error type in the instructor reports. These labels aid instructors in reading the reports more quickly. Rather than seeing that 30% of students responded '17', a label might inform the instructor that these students added when they should have multiplied. Labels should be aligned with the feedback given.

Each response condition may specify one or more criterion. Each criterion references a specific input field in the prompt. Input fields may be evaluated separately, together, or in any combination. For example, if two numbers must be in agreement they can be evaluated together. This allows for complex questions with more than one input field or more than one part to be scored or more than one step in the problem. The many to many mapping from input fields to question parts allows for complex scoring rubrics.

Consider the table (**Figure 3**) below. It defines the response conditions for a question where students are asked to add the two fractions: 1/4 + 1/6. Response conditions are considered in the order defined. Response processing stops once a matching response condition is found. Therefore response conditions are listed from most specific to least specific.

In this example, a correct response of 5/12 earns 10 points and receives appropriate feedback. The response is condition is labeled 'Correct' to clearly indicate in reports that this is the desired outcome. The next two conditions target common student errors. In the first, if the response is not expressed in lowest terms, partial credit is awarded. The response condition groups responses of particular error type and provides targeted feedback to students and identifies the frequency of this mistake for instructors. The remaining response conditions match increasingly more general types of errors until bottoming out at "any response" and "no response".

Response Condition : Correct	Score = 10
criteria: numerator = 5 AND denominator = 12	
feedback: Correct!	
Response Condition : Correct, not simplified	Score = 5
label: Correct, not simplified	

criteria:	numerator = 10 AND denominator = 24	
feedback:	Although this answer is equivalent to the correct answer, it is not in reduced form…	
Response Condition : Multiplied instead of added		Score = 0
criteria:	numerator = 1 AND denominator = 24	
feedback:	You have multiplied the fractions as opposed to adding them.	
Response Condition : Denominator correct		Score = 0
criteria:	denominator = 12	
feedback:	The denominator is correct and in reduced form, but the numerator is not correct.	
Response Condition : Numerator correct		Score = 0
criteria:	numerator = 5	
feedback:	The numerator is correct and in reduced form, but the denominator is not.	
Response Condition : Incorrect, other		Score = 0
criteria:	<any response>	
feedback:	Incorrect.	
Response Condition : No response		Score = 0
criteria:	<no response>	
feedback:	You should attempt to answer the question.	

Figure 3: Example response conditions for a question asking students to add two fractions.

Each question type (**Figure 2**) supports different types of response criteria.

Criteria for multiple-choice questions are specified by describing the set of choices selected by the student (choice {A} or choices {B, C}). An exact set of choices may be given or a set theory operator can be used. For example, one can check if the student's response is a subset or superset of the correct response. Multiple-choice multiple-select questions, while being bound in terms of student input, require flexible criteria such as subset (student missed some correct choices), superset (student selected more than correct choices), etc..

Open ended questions (text, number) add further need for rich responses, particularly when one question may contain multiple such inputs. In numeric questions like the one above response criteria may be defined as an exact number to match ("37"), an inequality (">= 15"), or a range ("[4.5, 15]"). Each criterion may also require a specific number of significant figures. In text questions exact strings or regular expressions are used.

For ordering questions, an exact or partial order may be given. It is also possible to compare the relative positions of the items as ordered by the student. Consider a question which asks the student to order audio clips as they might occur in a conversation between two speakers. Initial feedback may encourage the student to place the greeting first. Subsequent feedback could check to see if a question occurs before its response. Further levels of targeted feedback may then be authored, supporting the student until the correct answer is reached.

Although our emphasis for this paper is on feedback provided to students and instructors, the assessment system also provides a basis for generating more traditional exams, and so supports flexible scoring practices as well. The assessment system supports both partially correct and partially incorrect responses. The score assigned determines the degree to which a question is correct or incorrect. A correctness threshold determines the minimum score required for a question to be considered correct. For example, if a question has a correct threshold of 5 and the student earns greater than or equal to 5 points then the question would be marked correct. If the student earns less than 5 points then the question would be marked incorrect.

Thus, a student score may be correct but not achieve full points on a question if a question was authored in a way to allow this. Additionally, point values of questions do not need to bound the actual score earned by the student responses. Answers that are worth above and beyond a base point value of the question (bonus points) or below zero (penalty points) are also possible. For example if the question is worth or "out of" 10 points, a response condition might assign 12 points to indicate a fully correct answer and 2 additional bonus points.

Last, the assessment system supports pools of like questions to support student practice should their experience encourage them to keep trying new questions. Pools also provide a convenient way to group like questions, useful for organizing the authoring and reporting processes. Pools encourage course designers to think in terms of learning objectives. If each pool is intended to measure a single learning objective, learning activities and quizzes can be assembled easily for any unit of content by drawing questions from the appropriate pools. Accordingly, feedback reports to the instructor can then summarize student performance across like questions.

Flexible Feedback Examples

In **Figure 4**, a question is posed that requires the student to add two fractions together and give their answer in simplified form. The author of the problem knows a few common mistakes students make when answering the question. Students may reach the correct value but not simplify the result, or they may reduce the denominator but not reduce the numerator correctly. Each of these (and other) cases are identified by the author and appropriate feedback is given for each. In this example, the student correctly added the fractions but has not reduced the result. The system matches this case to the appropriate response condition written by the author based on the previously discussed criteria, and the student is given feedback tailored to their common error.

Figure 4: An example of constructing an assessment question so that the student is given feedback tailored to their common error.

Just as importantly, the author labels these mistakes so that when the instructor views the performance of their class, rather than simply seeing all of the values entered for the numerator and all of the values entered for the denominator with a frequency, the instructor receives feedback about which *categories* of mistakes student made **Figure 5**.
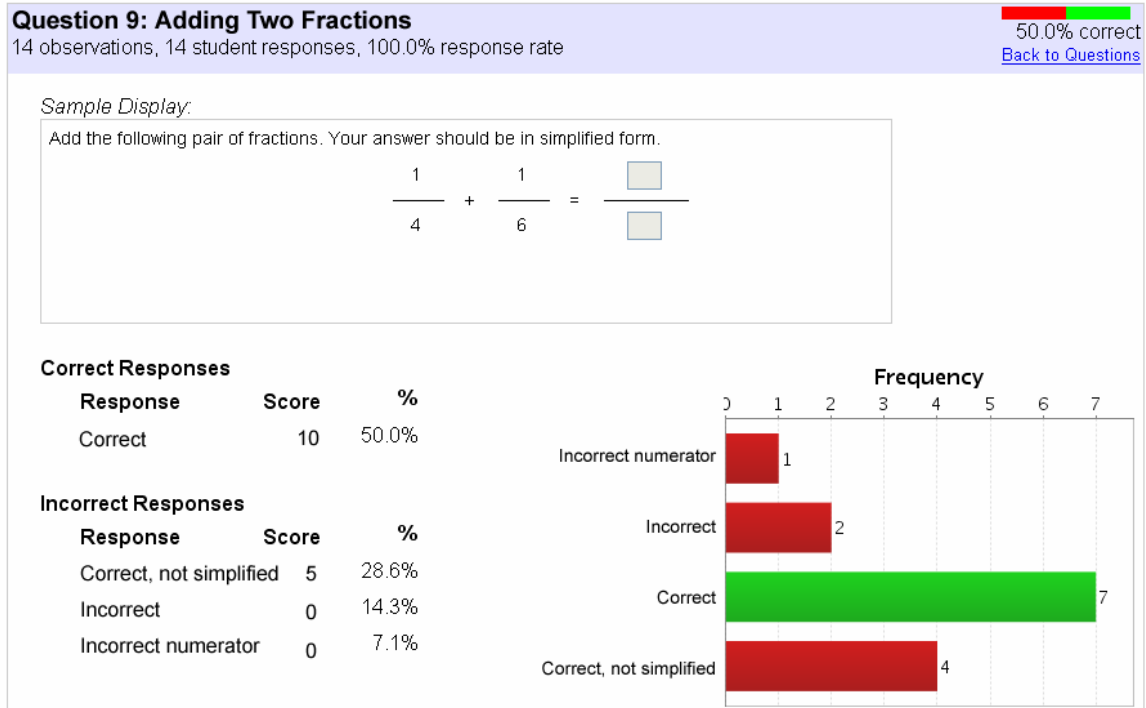
Figure 5: An example of constructing an assessment question so that the instructor sees responses broken down into response categories rather than raw inputs.

Ordering Example

In this example**,** the student as asked to sort phrases from a two person conversation in French. Note that this example uses text for the purposes of presentation in this paper. A real exam may include audio clips in place of text. For this example we assume a common error to be the placement of "Pourquoi?". In **Figure 6**, the student has swapped the order of the last two lines as displayed. The response processing provides specific feedback on this error, as we see in the figure, if ever the line beginning "Parce qu'.." is ordered prior to the asking of "Pourquoi?". As in the previous example, the instructor report would categorize this response with any other students who fell into the same mistake, regardless of where in relation to the rest of the conversation they make the error, simplifying the view of responses greatly. Many different distinct overall orderings may be entered by students, but those responses can still be grouped into the pre-defined response conditions.

Figure 6: another example of targeting feedback to illustrate a common mistake.

Future Work

Some amount of attention is planned to expand features of the assessment software system going forward. Interface work enhancing "practice" styles of interaction is planned, as well as the possibility of allowing even more flexibility in question generation, such as parameterized question templates.

Using the flexibility designed into our new assessment software system content authors are beginning to design activities in ways that provide immediate targeted feedback to students, instructors, and collect meaningful data for our upcoming Digital Dashboard for Learning (DDL) (Bajzek, Brown, Lovett, Rule, 2007) project. The DDL is an instructional tool that collects learning data as students work through multiple online instructional activities (not just assessment), summarizes the data, and displays meaningful indicators of students' progress so that (1) faculty can monitor students' progress and intervene where necessary and (2) students can monitor and adapt their own progress through their learning. This tool is one focus of our ongoing research activity.

For the future, the DDL might compare students reported confidence against their actual performance. By combining individual feedback tools, greater depth of information is available. In constructing the DDL it will be important to identify ways to bring together currently discrete sources of learning data such as tutoring systems, assessment questions, instructor reports, etc., into a common display to support learners and instructors as they progress through their online learning experience.

References

Bajzek, D., Brown, W., Lovett, M. & Rule, G. (2007). Inventing the Digital Dashboard for Learning. In C. Montgomerie & J. Seale (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007* (pp. 1084-1092). Chesapeake, VA: AACE.

Lovett, M. C., Meyer, O., & Thille, C. (in press). Open Learning Initiative: Testing the accelerated learning hypothesis in Statistics. *Journal of Interactive Media in Eduction.*

National Research Council. 1996. *National Science Education Standards.* Washington, DC: National Academy Press.

National Research Council. (2004). *How People Learn: Brain, Mind, Experience, and School.* Expanded Edition. Washington, DC: National Academy Press.