Beyond Multiple Choice: Computer-Mediated Practices and Assessments to Support Higher-Order Objectives

©Elizabeth Dalton
October 2001, May 2002

Overview:

A variety of factors are contributing to the increasing use of computer-mediated assessments. First, there is an increasing aware of the need for assessments themselves. Research and experience have taught us that instruction must include practices and assessments, as well as presentations, or there is little or no value-add over the provision of document collections. As training budgets become tight, there is a need to compress training time, to provide "just enough, just in time" training solutions, which need to be guided by assessments of some sort. There is also an increasing awareness of the need to be able to evaluate and quantify the benefits of training. Assessment forms one of the four recognized levels of training evaluation (level 2, "Learning," according to Kirkpatrick).

Assessments can be costly and time-consuming to administer and score. One way to reduce this expense is to automate as much of the process as possible, whether the assessments are being administered as part of an local instructor-led or remote, asynchronous technology-based event. There has also been a marked increase in the need for distance learning solutions (due to expense restrictions, safety concerns, and a need to avoid time out of office and away from families). This has often led to a call for technology-based instruction, for which technology-mediated assessments are indicated.

Yet computer-mediated assessments, as commonly implemented (mulitple-choice or limited text matching), measure very low-order learning objectives. Most are restricted to simple recall of facts and definitions. This presents a critical problem: while low-order learning objectives meet some needs for quick recall and form the building blocks for higher-order objectives, they do not enable far-transfer of skills from the learning environment to the performance environment.

What is needed are mechanisms which will allow the automated or mostly-automated administration and scoring of assessment items and sets which can measure higher-order learning objectives. This whitepaper attempts to describe such assessment forms.

Bloom's Taxonomy and the Value Pyramid

Benjamin Bloom and others have contributed a useful "taxonomy" by which educational objectives can be categorized, which has become the standard by which cognitive objectives are evaluated. The six levels of Bloom's taxonomy are listed below, from highest-order to lowest. Next to this list is Christopher Barnatt's data progression value pyramid.

6.0 Evaluation

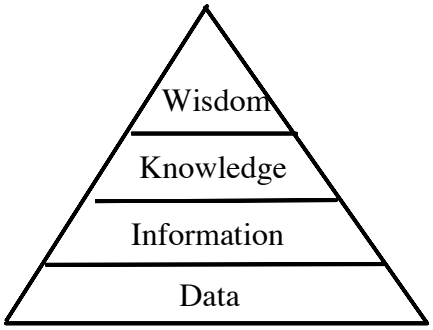
5.0 Synthesis

4.0 Analysis

3.0 Application

2.0 Comprehension

1.0 Knowledge



The pyramid shows how the volume of available data which can be gathered about a given subject must be gradually refined, concentrated, and structured to provide a high value to the learner. At the lowest level, one can simply provide the learners with access to data, without distinguishing the value of one element from another. This provides no instructional value-add. At the "information" level, someone with some expertise in the subject has sifted through the data and provided judgments about which data elements are worth study, but no further building upon those elements has occurred. This level of the pyramid might be thought to equate to Bloom's level 1. At the "knowledge" level (not to be confused with Bloom's use of the word "knowledge"), the learner is expected to be able to do something useful with the results of their learning experience. This might be thought to be similar to levels 2 and 3 of the Bloom's taxonomy. But the ability to perform independent, original analysis and synthesis is more clearly associated with the highest point, "wisdom", and only at this point do we expect learners to be able to make valuable judgments and critical decisions in a subject area.

A considerable amount of effort must be expended to sift "wisdom" out of "data", and a similar effort is necessary to determine whether the learner has achieved the corresponding learning goals. Yet this is not an insurmountable task, and much of it can, in fact, be at least partially automated, as we will see. If we are prepared to specify high-order learning objectives, we have a variety of options available with which to assess them.

Objectives and Activities

The easiest way to become familiar with the different levels of Bloom's taxonomy is probably to see examples. Below are typical learning objectives, including the categories they belong to and examples of how they might be assessed. Descriptions of each of the specific activities follow the table. Many activities can be used for more than one taxonomy level. The difference is in how the assessment item is prepared and scored. These differences are discussed in some detail following the table.

Objectives	***Examples***	***Activities***
1.0 Knowledge: Facts	Memorize: Settings	Simple objective assessments (multiple choice, short-answer, matching) Simple games (concentration, crossword

Objectives	*Examples*	*Activities*
Concepts Principles Theories Categories	Commands Parameters Definitions	puzzle) Sorting exercises
2.0 Comprehension Translation Interpretation Extrapolation	Read code Read config files Read error messages	Simple interpretive exercises Discussions
3.0 Application Procedures & Processes	Install Configure	Diagnostic labs Remote labs Instrumented labs Simulations (install/config) Roleplay Discussions
4.0 Analysis Elements, Relationships, and Organizational Principles	Troubleshooting Error-finding Root-Cause Analysis Organization	Interpretive exercises - complex case studies Simulation (break/fix) Remote labs (break/fix) Sorting/Organization exercises Concept Maps Jigsaw Puzzles (flowcharts, organization charts, etc.) Discussions
5.0 Synthesis	Writing new code, specifications Customization Configuration design Planning Designing	Case studies in interpretive form Simulations (graphic or text) Discussions Peer review - slashdot Mentor review Scavenger hunts Concept Maps Sorting/Organization exercises Automated Essay Scoring
6.0 Evaluation	Configuration design Evaluating alternates	Peer review - slashdot Mentor review Case studies Scavenger hunts

Detailed Descriptions of Activities

These descriptions are intended to provide an overview of each of the activity types listed above, including how they can be prepared, when they are most useful, and advantages and disadvantages of each.

Simple objective assessments

These are simple multiple choice, short-answer, and matching assessment item types which are capable of testing learner memorization of a variety of facts. Many of them depend on a simple definition or image, paired with a short text item.

Simple Games

These are games such as "concentration" card games or crossword puzzles. They use the same data types as simple objective assessments, but are somewhat more engaging for the learners. "Quiz show" formats are often especially popular. Scores may be saved and displayed for competitive purposes, though there are risks to exposing learner identities in this way. The use of "aliases" or "screen names" can allow for the positive benefits of competition without the negative consequences of the loss of privacy.

Sorting Exercises

These are variations on matching items, allowing learners to practice or demonstrate ability to use predefined categories and definitions. They can often be implemented as games. "Arcade" type games may be used when quick recall of simple facts is important. For higher-order Analysis and Synthesis objectives, learners may be allowed to construct their own categories as part of the exercise. The level of complexity of these types of activities can approach that of simulations. Concept Maps may be regarded as a specific form of sorting/organizational exercise. Scoring concept maps and similar learner constructions may be difficult in an automated environment, but see "Peer Review" below.

Interpretive Exercises

These are useful for objectives which involve interpreting code or error messages. A sample of data is required (e.g. source code, error output) and objective questions are asked about it. At the "Comprehension" level, the questions need merely involve interpretation, translation, or extrapolation. At the analysis level, the questions may address possible causes, proposed solutions, or further lines of inquiry. This is a flexible technique which can provide many of the benefits of simulations at considerably lower cost to prepare.

Discussions

Many kinds of learning objectives can be practiced and assessed in the form of discussions. Distance and asynchronous discussions are available in web-based "bulletin board" systems, email aliases, chats, etc. Scoring discussions beyond participation is difficult. (See "Peer Review", below.) Role-play discussions, in particular, are useful for practicing and assessing the application of soft skills.

Labs

Lab exercises can be automatically administered and/or scored in a variety of ways. Many techniques described here can be used simultaneously. In "diagnostic" labs, the automated assessment system does not have direct visibility to the lab activity, but asks the learner a series of questions designed to allow the system to determine how the learner has performed on the lab, e.g. by asking the learner to select from a set of possible response messages during an installation. In this way, diagnostic labs resemble multiple-choice questions, but the ability to customize feedback per distractor, rather than merely marking one or more distractors as "correct," is critical. Diagnostic labs can be especially helpful in the practice portion of delivery, but can be used during assessment, as well. Remote labs allow physical hardware to be centralized, but still enable the learner to perform lab activities on actual systems via network connections. These are suitable for software topics. Automated services can be used to prepare the lab environment so that each student has a consistent learning experience. "Instrumented" labs are those in which the automated assessment system has some direct visibility to the environment being used in the lab, as in database monitoring, hardware sensors, or other monitoring processes. These can be difficult to construct, but offer the most powerful forms of guidance and assessment of all automated labs. Labs are usually used to allow the learner to practice and perform Application and Analysis learning objectives.

Simulations

Simulations do not duplicate the real environment in which the performance would take place, but provide enough elements that the learner can practice and demonstrate their skills. Simulations may be used for installation and configuration activities as well as troubleshooting exercises. Simulations can be expensive to create, but sometimes the "real" environment would be even more so. Simulations may provide Application level practice and assessment by simulating installation and simple configuration activities, or may be used to practice and test Analysis and even Synthesis objectives by adding more complex situations. Simulations may be based on text interfaces or GUIs. A key advantage of simulations over real-world practices is their ability to offer enhanced feedback and guidance during the experience.

Jigsaw Puzzles

Many concepts, such as flow charts, organization charts, and other relationship topics, are most easily understood in graphic form. They can be tested in this form as well, using "drag & drop" interfaces. Scoring mechanisms evaluate the positional relationship between the draggable objects.

Peer Review

Peer Review (and the enhanced version, Mentor Review), are scoring mechanisms for complex submissions, rather than being activity types in and of themselves. For many learning objectives, a large body of examples is helpful. There is also a value in allowing learners to participate in constructing and providing examples. However, the issue of evaluating these examples, either for inclusion as part of a course or as a formal assessment mechanism, remains. Recently the technique of "peer review" has been popularized by online services such as SlashDot and Amazon. In these services, anyone

can submit an item, and then other members of the online community are invited to comment on the value or quality of the item. Over time, online "reputations" of items, contributors, and evaluators develops and these characterizations can be used to sort and prioritize contributions. While the number of contributions to any individual learning objective activity might remain somewhat small, the ongoing contributions of a given contributor across all learning events in which they participate might allow for this kind of service to function well.

Mentor Review

Mentor review expands upon peer review by adding an authoritative status to some reviewers. Not all mentors need be members of the training delivery or development staff, but a core of dedicated mentors must be supplied to ensure that all contributions are scored and appropriate feedback is provided to learners in a timely manner. In cases where scores will be tracked and used for other purposes, a scoring rubric must be provided to each mentor, and mentors themselves must be trained and assessed in their ability to use the rubric. Notifications or workflow management tools must be provided to ensure timely evaluation of submissions.

Scavenger Hunts

Many topics benefit from the inclusion of a broad array of external sources, with the world-wide web being one of the most extreme examples of such a source. In some cases, the instructional goal might actually be increased familiarity and expertise with a resource. In a scavenger hunt activity, learners are directed to collect links to specific external sources and documents. Scoring such contributions can be done by peer review, mentor review, or in some cases, by self-scoring or self-assessing. If peer review formats are used, collected references can form a learner-developed enrichment to the learning materials (possibly with mentor approval).

Automated Essay Scoring

For learning objectives which involve written text, new techniques allow automated scoring of assessments. These do require either a large body of scored examples (possibly growing from a peer review or mentor review process), or a coded scoring rubric.

Guided Self-Assessment

Learners may be assisted in guiding their own assessment process. By asking careful diagnostic questions of learners, the assessment system can allow learners to be active participants in the scoring process, exercising their own human judgment. Guided self-assessment takes more time to construct than subjective self-assessment, but the results are more reliable. Effectively, the learner is being trained in self-assessment as part of the learning event process.

Subjective Self-Assessment

When no other assessment mechanism is available or appropriate for a given learning objective, the learner may be asked to self-assess their performance based on specific objectives. This method is the least reliable of all assessment methods discussed here, but is simple and inexpensive to prepare, and is better than no assessment at all.

Conclusions

While implementing some of the activities above would require the development or acquisition of new tools, many can be implemented using tools available today within the IT education environment. Interpretive exercises, in particular, show promise as a low-cost way of assessing higher order objectives. Further analysis needs to be done on other options, such as peer review systems, to determine whether they are currently feasible to implement. This analysis is currently in progress.

References:

Barnatt, C. (1997). *Challenging Reality. In Search of the Future Organization*. Chichester, England: Wiley.

Bloom, B.S. and Krathwohl, D.R. (ed.) (1989) *Taxonomy of Educational Objectives Book 1: Cognitive Domain*. New York: Longman Publishing Group

Linn, R.L., Gronlund et al. (1999) *Measurement & Assessment in Teaching (8th ed)*. New Jersey, Columbus: Merrill Prentice Hall

Shann, M.H. (2001) *Performance Assessments for Practitioners*. (In Press)

Sharpe, C. (ed.) *The Info-line Guide to Training Evaluation*. Alexandria: ASTD

E-rater as used by ETS: http://www.etstechnologies.com/scoringtech-erater.htm

Lexical Semantic Analysis @ CU Boulder: http://lsa.colorado.edu/

Knowledge Analysis Technologies (Intelligent Essay Assessor): http://www.knowledge-technologies.com/

Quia Web Learning Activities: http://www.quia.com/web/

SlashDot news service: http://slashdot.org/

Amazon Online Bookstores: http://www.amazon.com