

MCTE 628, Instructional Systems Design

Portfolio #1-4: Chapter 4, Assessing Learning

Exercise A: *Identify What Is Being Assessed*

Exercise C: *Recognizing Ways to Achieve Greater Reliability in a Test*

Exercise D: *Recognizing Test Items Likely to Lend Greater Validity to a Test*

by Leanne C. Boyd

Course: MCTE 628 — Instructional Systems Design
Subject: Assignment #1-4, Chapter 4, Assessing Learning
Professor: George Fornshell, Ph.D.
Student: Leanne C. Boyd
Usercode: boydl
Email: boydl@scis.acast.nova.edu
Due date: February 14, 1999





Determining What Is Being Assessed

One feature that will stand out in a study of the action of assessment is that "research documents that tests are learning events." (*Seels & Glasgow, 1998, p. 83*). It has been my experience that, unfortunately, the student rarely sees it this way! The axiom and hope of this course of study is that each student will search out an individual method for applying the highest level of accuracy and usability in their instructional design environments. My primary discovery in this chapter, *Assessing Learning*, was that there were very good reasons, as a student myself, for NOT appreciating many forms of testing that habitually have been applied to a learning curve that, otherwise, was quite gratifying. Since young ages, we have been subjected to what has been nicknamed as "drill and kill." It is this writer's hope that, as we watch the very foundation of instructional design undergo massive changes by the very nature of the Internet and all of its possibilities, that there will soon arrive a new mind-set concerning ... assessment.

Since up to this time, there has been no perfect mode of assessment for learners, we all must live with the fact that most of the time, a varying number of methods must be considered in projects. Instructional design faces the same challenges as any other entity, in the face of the rapid changes that are happening in digital learning arenas. It must be asked: If there were a perfect and all-encompassing method for assessment, and the major problem with this particular method was that it was very labor intensive and therefore costly, would a suitable argument for using the method be that the increased gains would more than offset the increased costs? (*Wiggins, 1990, p. 1, as cited in Seels & Glasgow, 1998, Pp. 88-89*).

I strongly feel that this chapter was incomplete in its overview of assessment, as it did not address one issue concerning the impact of digital and global descriptions, on the act of assessment. We have entered a new realm, and it is my contention that this has created the necessity for querying every aspect of building and presenting "instruction." A by-product of this is that new ways of looking at the assessment process must also be re-evaluated.

In comparing things learned in the study of this chapter, I am providing a list of a few of the instructional design projects with which I have been involved:

- 1) Math application for 4th-graders, which was available both via Web-based technology, as well as CD-ROM. It included games, animations, video, and many interactive elements. It also included automatic assessment of student activities, providing teachers with instant, printable assessment results.
- 2) Two interactive "research manual"-type software concerning artists involved in the international space race, and the Hopi Tribe of the Southwestern U.S. Both were constructed for use by kiosk, CD-ROM, and the Web.

- 3) Curriculum and course design for online college courses, utilizing existing documentation. All projects were constructed in HTML, and included hypertext as well as many multimedia elements.
- 4) Task-based "user's manual" for technicians worldwide, in the Lucent Technologies communications' equipment divisions. The project was entirely Web-based;
- 5) Interactive "city guide" that instructed users concerning events, locations, and diversions in the Denver, Colorado area. The project was a kiosk, but could be transformed into a Web-based application.

This list was to show not only a variety of instructional design products, created with a wide variety of tools and methods, but also to indicate that — since 1995 — not one project was outside the realm of the use of the Internet and World Wide Web. Each learning structure, and all built-in offerings for assessment, was literally "outside" the traditional descriptors for "assessment." This is centered in one word: hyperactive content. This is an arena where the rules for assessment *change*. The reason for this is that, in systems with enormous amounts of information, with assorted formats, one major evaluation problem would be the notation of *what is accessed*. This could be true even in a paper-based library system, as well as any digital system. However, in the emerging electronic systems that provide high levels of learner control, THE major evaluation problem is calculating how learners conduct this control and freedom, as they are learning. The next scenario is in the dynamic, collaborative systems, where the notable assessment enigma is *who does what* — as well as *how it is done*. (Jonassen & Mandl, 1989, p. 360).

Another shortcoming I saw in the chapter was that not much was presented concerning assessment and its relationship with STANDARDS. I am speaking of basically what has come to be known as standardized curriculum and the emphasis that has been placed in recent years, on districts and teachers, for providing useful methods for comparison on state or even national levels. This was a very "hot" issue in the Denver District, in the early- to mid-1990s. During the summer of 1993, I worked with a group of teachers and administrators on re-developing high school curriculum items and issues. The project was part of a Perkins Grant; the team was extremely hard-working, knowledgeable experts, with a genuine desire to complete a prototype for use within the state, for providing all teachers a tool for assessing high school subject areas. The primary question, in the beginning was: Where To Start?

It must be pointed out that, in many ways, this project was doomed to fail. Much of the work was never utilized — not at the state level — nor even within this particular high school. In researching for this writing, I found documentation that provided insight into WHY the project wasn't as successful as it might have been. The Mid-continent Regional Educational Laboratory (McREL) states that the process might begin with using the "standards document" produced by one's own state. They show that every state except one is developing or has developed state *standards*. A truly disturbing research fact is:

"[A] study conducted by the American Federation of Teachers (AFT) reports that only 13 of 49 state documents are specific enough to be used effectively by teachers. The majority of state documents describe standards at levels of generality that do not provide sufficient clarity for classroom instruction, nor are they precise enough to serve as an instrument of accountability." (*Marzano, 1997*).

My group, in 1993 wished — as all schools and districts wish — to adopt valid standards. Furthermore, *parents* want to know how their children are performing academically — compared with standards. The next suggestion by McREL was to use the national "standards documents" such as the ones published by national associations and subject matter groups. Unfortunately, McREL points out ... "these documents commonly embed a description of requisite knowledge and skills within lengthy descriptions [over 2000 pages and 14 pounds of documents] of performance activities, curriculum goals, instructional strategies, and the like." (*Marzano, 1997*). As pointed out by our authors, the assessment process is very difficult. In 1993, I viewed much of the process as either the act of "reinventing the wheel," or, a commitment to NOT be doing so.

For all of my lifetime, my experiences as a student have indicated to me that most assessment FEELS like the process is undeniably tied to the act of GRADING. Seels and Glasgow made the point that "when we assess, we estimate or judge the value of a person, activity, or situation. When we approach assessment from this point of view, it becomes easier to understand that assessing learning is not equivalent to grading. While one function of assessment can be to provide a basis for grading, that is not its primary function." (*Seels & Glasgow, 1998, p. 82*). This, to me, was an enlightening statement. If this is truly the issue — and, I feel that it *should* be — then, why does assessment almost always have the competitive overtones of ... GRADING?



Assessment And The Missing Element: Psychological Factors

This, then, is a third element that wasn't explored in our chapter. For the student, and even for the instructional designer or the administering teacher, the assessment process is tied closely to psychological factors. My research led me to a new concept (for me). This concept is called "*knowledge structure measures*," which are based on and enriched by psychological theories. The postulation is that information in one's memory is organized or "structured" so that it facilitates the storage, recall, and manipulation or handling of information. Much research shows that "assessment of knowledge structures may be used to reliably distinguish skilled and unskilled performers in a variety of domains." (*Rowe et al., 1997*). This indicates that this form of assessment is usable in a variety of settings. It also strongly shows that perhaps one very powerful missing element in traditional assessment methods may be the internal, psychological aspects of learning and evaluating the individual grasp of information. I propose that it is because the traditional example disregards psychological factors, and instead, offers ... "drill and kill."

The interesting difference in this example is that *measures of knowledge structures* may be used to advance toward ... and evaluate ... *conceptual* understanding for assessment and training purposes. In a typical situation, the value of an individual's knowledge structure is decided by comparing it to a *standard knowledge structure*. This would be a collection of the structures of several experts. Ideally, this approach indicates a positive *linear relationship* between the progress toward expert task performance levels, and the development of *knowledge*. Each learner is given a "knowledge score" by calculating the comparison of his or her knowledge structure against the standard knowledge structure. This description sounded very good and reasonable, up to this point. Then my research source stated that "a validated knowledge structure measure may then be used to extract rich cognitive representations that are useful for assessing and diagnosing student knowledge and identifying targets for training intervention. **This knowledge score may then be validated by determining how well it correlates with a recognized performance measure, such as final exam or final course grade.**" (Rowe et al., 1997).

With great dismay, this had led back to ... grading. The search for an elevated form of assessment ground to a halt.

This ultimately sounded much like our textbook's description of CRTs (criterion-referenced tests), in which the use of the term "criterion" refers to the affinity between the objective and the method of assessment, and to the level of performance needed for mastery of a skill or subject. The CRT score gives information about a learner's mastery of one specific skill relative to the objective, and therefore shows one's ability in performing specific competencies. (Seels & Glasgow, 1998, p. 83). A CRT score lets everyone know exactly how well the student stands, relative to a *standard*. (p. 84). The workplace is becoming more technologically complex. It demands increasingly specialized skills, workers, as well as students who will be entering the workplace, must rapidly become skilled in a broad range of *cognitively* complex tasks. The cognitive aspect of a task must be spoken to, for purposes of *meaningful* assessment. Since the traditional method of assessing the cognitive domain is via a paper and pencil test (p. 83), are we then still stuck with *grading* and "drill and kill" methods?



Greater Reliability And Validity in Tests

Reliability means that a test will render a dependable measure, so that if a test is repeated the same results will be obtained. Reliable tests have consistency and temporal dependability (a student will show the same level of mastery on one day of the week as on another day). Reliability is augmented by an architecture with items that are not ambiguous, and by scoring the test as objectively as possible. (p. 85). Validity means that a test will measure what it is *supposed* to measure. Without reliability, there can be no validity. (p. 85). The best approach for substantiating a test's validity is to provisionally demonstrate that the test accurately discerns the masters from the non-masters of the test items. This is to say, items that do not make this distinction, should be revised or completely dropped from the test. (p. 86).

Using guidelines from the text and outside resources, one can then come up with an improved understanding of assessment procedures as a historical process, as well as perhaps mapping out an "ideal" evaluation plan. It appears that using a variety of frequent assessment techniques in the classroom may be a preferred approach. According to McREL, in the traditional classroom, the teacher is responsible for assessing students on standards and benchmarks. They are therefore able to utilize a variety of techniques, including portfolios, performance tasks, traditional tests, and natural observation. There is a decided advantage to this approach, in that assessment of standards and benchmarks does not diminish regular classroom instruction. The assessment process is integrated into the regular routine of the classroom. One major risk with this method is that, unless the assessment agenda is constructed and planned carefully, there is no guarantee that teacher evaluations will be reliable and valid. (Marzano, 1997).



A Possible Solution — Authentic Assessment

This was further corroborated in my research. . David Jonassen and Heinz Mandl, editors of Designing Hypermedia for Learning, say that "if evaluation is to address the processes as well as the products of learning with a new technology, a *variety of methods* must be used ... the solution to the problems of evaluation is to take *a multi-faceted approach to evaluation*." (Jonassen & Mandl, 1989, p. 360).

It is my opinion that the world of hypermedia lends one very important element to not only enhanced learning, but also to the assessment process. That is that by sheer use of the computer to study, research, and learn — the student has literally placed herself on an equal level with her workplace counterpart. The computer is the tool that performs real tasks, both in learning and in performing work in the work world. It is somewhat different than the equal use of paper and pencil, or research libraries, or any other common element. It is, more than any tool in prior history — an extension of our selves. The assistance of the computer quickly becomes a cybernetic "right arm" for the student or the worker — an indivisible part of interactivity with the real world, an extension that increases meaning in our lives, as much as our hands or our sensory organs. It is almost as if the computer, with all its offerings for learning, is part of what must be assessed. Jonassen and Mandl express that because of the personalized nature of learning (and being assessed for that learning), it should be realized that "the essential problem of evaluating highly interactive systems is in measuring both the quality of the interaction as well as the product of learning. Evaluations of hypermedia-based learning must address both the process of learning and the outcomes of learning. (p. 360). I see that process of learning also cannot be divided from the avenue for processing ... which is the computer.

Our textbook examines one assessment form that perhaps is the closest to the Jonassen and Mandl proposal. That is *authentic assessment*. The concept of authenticity in assessment means that performance is contemplated directly in proportion with "real" tasks. The opposite of authentic assessment is the "drill and kill," paper and pencil test

(Wiggins, 1990, p. 1, as cited in Seels & Glasgow, 1998, Pp. 88-89), from which learning must be deduced, and from which so very much about our academic and professional lives, are determined. More than ever, the new Age of Technology calls upon all of us, as students, and ultimately, as professionals, to perform what amounts to authentic assessment. Simply because we now are forced to bring the real world tasks of lifelong learning into the workplace, as well as expect that students will increasingly perform according to professional standards — it is my vision that some form of authentic assessment will BE that ideal mode of evaluation that we seek.

If nothing else, as designers, we must look for "test validity" that shows the test to be a simulation of real-world analyses of ability. "Authentic tasks involve "ill-structured" challenges and roles that help students rehearse for the complex ambiguities of the "game" of adult and professional life. Traditional tests are more like drills, assessing static and too-often arbitrarily discrete or simplistic elements of those activities." (Wiggins, 1990, p. 1, as cited in Seels & Glasgow, 1998, Pp. 88-89).



Conclusion

The highly interactive media that is emerging challenges the evaluation of instructional products, as well as the learning that happens because of these products. This interactivity provides the learner with access to untold and diverse information. It gives the user a personal control over the process of learning, and it carries a vast potential for collaboration with not only the delivery system, but with other people. This tremendous empowerment of *the learner* forces evaluators of learning to embrace a broad-based set of methods and criteria. ONLY this will accommodate self-directed learning. (Jonassen & Mandl, 1989, p. 368). To leave this out of the assessment process would be to cripple the effort for advancing superior techniques in evaluation.

I see extremely dire consequences if we, as designers, do not heed this trend in learning evaluation. The Mid-continent Regional Educational Laboratory (McREL) states it this way:

"Do *not* hold students accountable for specific levels of performance on any standards. *This is the system we currently have in place.*

"In virtually every state, the only standard students must meet to graduate is that they obtain a certain number of "credits," and a credit is earned by obtaining at least a "D" in a course. The advantage of this approach is that it is very easy to earn a high school diploma. The disadvantage of this approach is that students can graduate without acquiring any specific skills and abilities, such as the abilities to read, write and compute." (Marzano, 1997).

For years, private and public sectors have cried out that "Johnnie cannot read," and they have asked, "Why?" It has been the purpose of this writing to show that the product of the

appropriate — perfect — assessment process will have an onward, spiraling effect on learners. Johnnie will not only *learn* more effectively, but he (or she!) will perform at much higher levels during evaluation.



REFERENCES

- Jonassen, David H., and Mandl, Heinz.** (Editors). (1989). Designing hypermedia for learning. *ASI Series F: Computer and Systems Sciences, Vol. 67*. Proceedings of the NATO Advanced Research Workshop on *Designing Hypertext/Hypermedia for Learning*, held in Rottenburg/Neckar, FRG, July 3-8, 1989. Assisted by: Fischer, Peter M. and Wang, Sherwood. New York, Berlin, Heidelberg: Springer-Verlag (in cooperation with NATO Scientific Affairs Division).
- Marzano, Robert J.** *Standards at McREL: Eight questions you should ask before implementing standards-based education at the local level*. The Mid-continent Regional Educational Laboratory (McREL): Aurora, CO. [<http://www.mcrel.org/about/>]. Modified: Sept. 24, 1997. [Online]. Retrieved Feb. 7, 1999 from the World Wide Web: <http://www.mcrel.org/standards/articles/8-questions.asp#6>
- Rowe, Anna L., Meyer, Thomas N., Miller, Todd M., and Steuck, Kurt.** (1997). *Assessing knowledge structures: Don't always call an expert*. Paper presented at the Annual Meeting of the Human Factors and Ergonomics Society, Albuquerque, NM, August 1997. [Online]. Retrieved Feb. 5, 1999 from the World Wide Web: <http://www.brooks.af.mil/AFRL/HEJ/fst/hfes/hfes97.htm>
- Seels, Barbara, and Glasgow, Zita.** (1998). Making instructional design decisions. 2nd ed. Ed.: Debra Stollenwerk. Upper Saddle River, NJ: Prentice-Hall, Inc.
- Wiggins, G.** (1990). *The case for authentic assessment* (Report No. EDO-TM-90-10). Washington, DC: ERIC Clearinghouse on Tests, Measurements and Evaluation, American Institutes of Research, p. 1. As cited in Seels & Glasgow, 1998, Pp. 88-89.
-