Running head: ONLINE ASSESSMENT

Online Assessment: A Study of the Development and Implementation of a Formative Online Diagnostic Tool in a College Developmental Mathematics Course

Tau Kadhi

Texas A&M University

Paper presented at the annual meeting of the Southwest Educational Research Association in Dallas, TX, February 5-7, 2004. Requests for reprints should be sent to Tau Kadhi, email: tkadhi@tamu.edu.

Abstract

This paper is addresses instrumentation design and an ongoing study of an online formative assessment instrument and its effects in pre-college (developmental/remedial) mathematics courses. The use of facet theory and instructional technologies are harmonized to construct and scale an online instrument designed to document student procedures while learning multiple meanings and models of fractions (MUL). Skills were coded a priori and mapped as per “facets of thinking” (Minstrell, 2002). Subsequently, this Fraction Diagnoser should provide a finer “grain size” than the Texas Higher Education Assessment (THEA) test and the THEA number subgroup (Levy, 1994).

Researchers focusing on assessment-centered learning environments contend that students do not always record the procedures they follow to attain their answers (Kulm, 1994; Lauvas, Haynes, & Raaheim, 2000; Romberg, 1992). Therefore, educational reform now stresses increased attention to “the process” as well as “the answer” (National Research Council, 1999; National Council of the Teachers of Mathematics, 2000). This process is a far more difficult one to implement because students are not required to list their thought processes or strategies on most assessments. Consequently, various researchers have suggested providing teachers with knowledge and tools to aid in deducing what approaches students have adapted (Hiebert, Carpenter, Fennema, Fuson, Wearne, Murray, 1997; Minstrell, 2002, 2001, 2000; Research Advisory Committee, 1998).

Computer-based assessment programs can assist in providing efficient and immediate feedback of student reasoning (Caverly & MacDonald, 2003). Also, informational and instructional technologies help to produce a personalized learning environment that can be learning goal specific with reasoning assessed immediately and in real time (MacDonald, 2001; National Council of the Teachers of Mathematics, 2000, National Research Council, 1999). For example, the computer program Diagnoser, designed based on the research done by Minstrell (2001), documents and evaluates students’ thinking as they solve physics problems.

Research Questions

This study will focus on the design, development, and implementation of an online formative assessment tool (henceforth referred to as the *Fraction Diagnoser*) that will identify facets for multiple representations and models of fractions (MUL). The

Fraction Diagnoser will be online program administered via the web to be used by individual students on a personal computer. Fraction Diagnoser will consists of sets of multiple choice items that assess student competencies in MUL. A limited number of items are set up in a corresponding branching structure and at the end of the sequence, a facet code will be assigned to the student representing path of the questioning taken. Eventually, the codes will be summarized to the facet the student has demonstrated while deciphering items responding to MUL.

Specifically, the overarching question that this research will focus on is:
How well can the Fraction Diagnoser assess students' depth/level of understanding of MUL as well as support instructional decision-making?

Developmental Questions:

- What are the facets in the fraction concepts and skills related to multiple meanings and models of fractions (MUL)?
- Does the Fraction Diagnoser identify distinct levels of understanding of MUL concepts and skills for individual students in developmental mathematics?

Implementation Questions:

- Do the data produced by Fraction Diagnoser provide information that can be used to describe student growth toward mastery of MUL?
- Can the data produced by Fraction Diagnoser aid in developmental mathematics teachers' instructional decision-making?

Participants

The participants will be approximately 100 pre-college level students and their instructors in various colleges and universities in east central Texas. The students will reflect a variety of cultural backgrounds, economic levels, and other key demographic factors. Teachers will be trained on the use of the Fraction Diagnoser and given the facet table. Interview and observational data will be collected to determine impact on instructional decisions.

Facet Theory

According to Shye (1978), “A classification of the item-domains of a given content universe according to some rule is called a facet of that universe” (p. 9), allowing broad usage of facets. Guttman, Minstrell, and Schumaker have all used facets or facet study in various research areas to study various hypotheses. Also, facet analysis allows for the collection of data both prospectively and retrospectively (Shye, 1978).

The theoretical framework design of Minstrell's (2001) *Diagnoser* provides a level of description and procedure for constructing models of student thinking that helps classroom teachers make instructional decisions and assist researchers by providing empirical data from which to make inferences. The empirical data comes from his theory of facets and facet clusters. Minstrell (2002) defines *Facet clusters* as “sets of related facets, grouped around explaining or interpreting a physical situation (e.g. Forces on Interacting Objects) or around some conceptual idea (e.g. Meaning of Average Velocity.)” (p.4). Also, facets are “used to describe students' thinking as it is seen or heard in the classroom or other learning situation. Facets of students' thinking are individual pieces, or constructions of a few pieces, of knowledge and/or strategies of reasoning” (Minstrell, 2002, p.4). To summarize, Minstrell has *facet clusters* (which are titled) composed of *facets* (which are numerically coded research-based observations).

Methodology

Development of Facets for MUL

The concept behind Guttman's Facet theory is to quantify and measure qualitative data. According to Levy (1994), Guttman's theory was originally designed to seek a facet design for mental abilities and eventually boiled down to seeking a definition for mental abilities. Guttman found that "difficulty with the old saw that 'intelligence is what an intelligence test measures' was that it was virtually facetless" (Levy, 1994, p. 513). Conclusively, Guttman's facet theory both defines the universe and scale it for certain types of data, citing the construction of structural hypothesis rather than with inference from samples (Guttman, 1958; Levy, 1994; Shye, 1978).

In addition to the other theories on Facets, Schumacker (1999) explains the usage of many-facet Rasch analysis. In his studies he explained the use of a crossed, nested, and mixed design of Rasch analysis to use for comparing facets. By using this method Schumacker (1999) offers the ability to create a vertical scale for each facet.

The facets for the Fraction Diagnoser will be developed by first drafting the initial mathematical goals for the instrumentation development. Those goals will be research based information concerning student understanding of fractions for students taking pre-college (developmental/remedial) level mathematics classes. The National Council of the Teachers of Mathematics (NCTM) Standards will be correlated with the American Association for the Advancement of Science (AAAS) Benchmarks and Texas Essential Knowledge and Skills (TEKS) to find specific learning goals for pre-college level students learning fractions.

Initial inspection of learning goals found three common problematic strands concerning fractions. The three major content strands for fraction were identified as Multiple Meanings and Models (MUL), Converting Forms (CON), and Comparing and Ordering (COM) (See AAAS/Project 2061). Unpacking of the MUL strand will lead to the development of the facets to be defined and scaled (Levy, 1994).

Unpacking the MUL learning goal will lead to a detailed checklist of competencies for mastery. In order to determine what research based objectives for MUL for pre-college mathematics students should be used, the TEKS, AAAS/Project 2061, Research Number Project, and NCTM standards were correlated. Then, research based items that test each competency will be listed. From both the objectives and the test items, the research based skills required to master MUL will be listed. Then, each skill or competency will act as a check for mastery and a specific code will be given for that assessment. Conclusively, an a priori MUL specific facet cluster, theoretically based from Minstrell's design, will be developed and used to document achievement and abilities and assess student understanding (see Table 1).

Table 1. Minstrell Facet cluster for Explaining falling bodies

*340 Gravitational pull by earth on falling object and mass of object compensate for each other. The resistance by the medium through which the object is falling increases with speed and will decrease the rate of acceleration

*341 (Fg – Fr)/ mass = acceleration (instantaneous rate) of fall. With no resistance, near the earth, things fall, accelerating at about 9.8 m/s/s.

342 Gravitational pull and mass compensate, but greater air resistance on the lighter object, making it fall behind.

343 Gravitational pull and mass compensate with no accounting for air resistance.

344 Greater drag effects compensate for greater gravitational pull explaining equal accelerations. No apparent accounting for inertial mass of the object.

345 Drag effects of medium will exist even when there is no motion relative to fluid medium. The resistive force exists even when the object is not yet falling.

346 All things fall with equal acceleration of about 10m/s/s.

347 All things fall equally fast regardless of medium effects. For example, vertical fall is at a constant velocity of 10 m/sec
348 Weight makes it hard to move things. The more weight, the slower they fall. It takes time to get them going. Heavier things will lag behind until they can get going.
349 Weight makes things fall. The more weight, the faster they fall.
349+ When you let things go, they fall
349++ Things fall down.

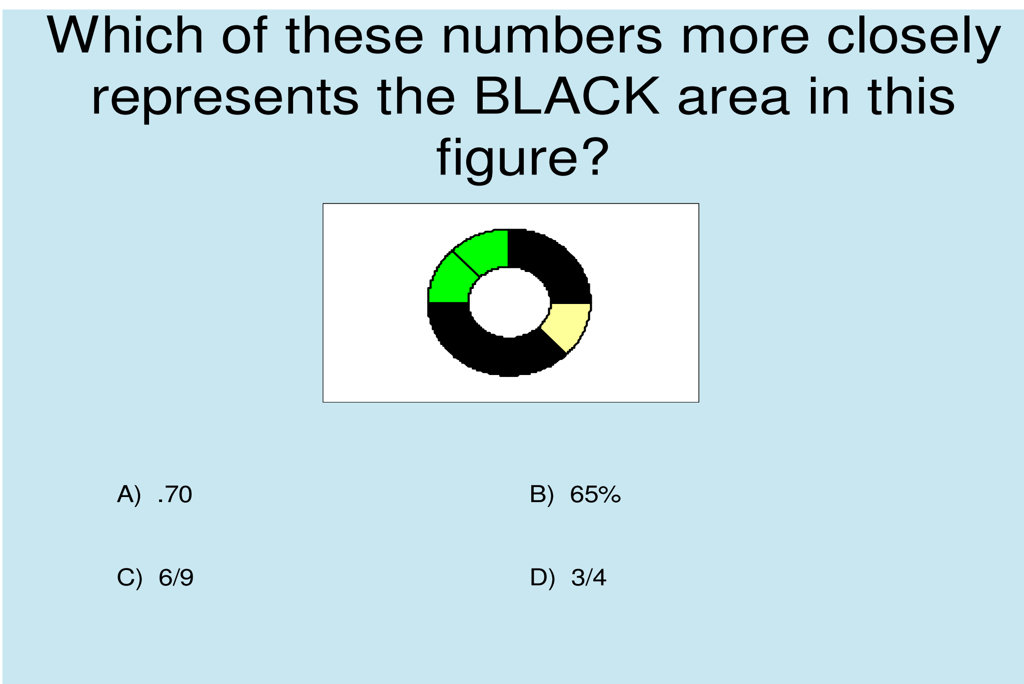
Note: For the cluster "Explaining Falling Bodies," the facet ending in 0 again represents a more conceptual understanding and the facet ending in 1 represents the more mathematical modeling of the situation. (Taken from Minstrell, 2002, p.22).

Design and Development of the Fraction Diagnoser

The Fraction Diagnoser program will be designed and developed following the Clements and Battista (2000) model for designing effective software. Initial learning goals will be drafted from national curriculum reform research and past research project studies. Then, an explicit model of student learning will be designed by using unpacked competencies from the learning goals. Next, the initial design for the Fraction Diagnoser will be designed using programs that are flexible an offer the ability for multiple representations of numerical information. Subsequently, Fraction Diagnoser will be piloted to investigate the components of the design, answering the questions of instrument feasibility and reliability. Finally, Fraction Diagnoser will be implemented in a full study, reevaluated, and rescaled. Upon its completion, the Fraction Diagnoser will be an online interactive assessment instrument designed to diagnose the understanding of specific concepts and skills about fractions giving detailed feedback of specific competencies for MUL.

The Fraction Diagnoser will consist of a series of multiple choice items written on many WebPages, linking questions based on responses. The multiple choice items used on the WebPages will be constructed using Minstrell's Diagnoser as a framework, where correct responses or distracters for each item will eventually link to a facet descriptor or component. For example, to choose a distracter from an anchor problem “a, b, or c” will link the user to another multiple choice problem that assesses specific sub-contents of the more complex anchor problem. If the student answers the anchor correctly (“d”), another problem of similar complexity will be given and eventually the student will be assessed according to responses.

A student will start with an “anchor item” for one of the three specific fraction concepts or skills. An anchor problem should begin with a general question that covers most of the components of the strand to be tested. These components were research-based skills for which there is evidence show mastery of the strand. For example, a MUL strand anchor problem could look like:



For example, the above anchor question asks that a student display their understanding of numbers, ways of representing numbers, relationships among numbers, and number systems; and whereas this question covers most of the mastery competencies when asked of understanding of multiple representations of fractions at the pre-college level, subsequent questions will completely cover the needed competencies.

The questions come in pairs. The first question typically asks, "What would happen if...?" Although it involves multiple choice answers, each choice is associated with one or another facet. Thus the system makes a preliminary diagnosis of specific, potential difficulty. The second question follows up with "What reasoning best justifies your answer?"(Minstrell, 2002, p.13).

Depending on the response to this anchor item, the student will be given subsequent items that are designed to define specific facets of the student's knowledge about the concept or skill. Since research based studies have shown that certain fundamentals are present when a student is able to perform these tasks (See Rational Number Project), test items in Fraction Diagnoser will be designed to eventually assess what fundamentals are present while attempting to master MUL.

Each student will respond to no more than eight and no less than five questions that cover all competencies on the checklist. After the students have responded to the questions and exited the program, Fraction Diagnoser will send the appropriate response path to a spreadsheet program for data analysis. The output data sent to the spreadsheet for participants for one administration, such as the pilot assessment, will look like:

	Q-1	Q-2	Q-3	Q-4	Q-5	Q-6	Q-7	Q-8
Student 1	1	1	3	4	3	0	0	0
Student 2	2	2	4	2	4	4	0	0
Student 3	2	3	2	2	3	3	2	0
Student 4	1	2	2	3	4	4	3	1
Student 5	4	3	3	2	2	2	1	4

Where the Q-1 column represents the students response to question 1 (the anchor), assigning numerical values to the multiple choice responses (a=1, b=2, etc.). Zero indicates that the student was not required to respond to question.

Although the instructor participant will be privy to the output data, this multiple classification of the content universe will produce a Cartesian product. This Cartesian will constitute a component set facet that will be summarized and scaled. The scale will be used to create the facet table that will be used to examine student abilities, document achievement, and assess progress.

Implementation of the Fractions Diagnoser

Students will complete the Diagnoser at three times: near the beginning, middle, and end of the units of instruction that develop the three fraction concepts and skills covered by the Diagnoser. In addition, the students Texas Higher Education Assessment (THEA) Test number subgroup test will be collected and correlational analysis of Fraction Diagnoser output will be run.

Data Sources and Analysis

The data for the study will be at three levels of "grain size." At the general level, student scores on the THEA will provide an overall assessment of students' achievement in developmental/remedial college mathematics. At the next level, the THEA number subgroup for number will provide evidence for student achievement on each of overall fraction knowledge and skills. At the finest grain size, the facet codes collected with the Fraction Diagnoser will provide additional, more detailed ordinal and descriptive data about levels of understanding on MUL.

The facet data will be compared with the THEA and THEA Number tests to determine the extent to which it predicts or explains variance on the tests. Patterns of student facet data will be examined to explore trajectories and patterns of knowledge and skill development. These results will be compared with Number pre and posttests to determine the value of the Fraction Diagnoser data in understanding how student mastery develops. Interview data will be analyzed to triangulate and confirm findings of the quantitative analyses.

Multivariate analyses will be employed to gain empirical evidence of the effectiveness of the instrument in providing valid and useful data. Multidimensional Scalogram Analysis permits the depiction of studied objects while making full use of the original data in its raw form while revealing relations that exist among them (Zvulun, 1978). Also, because of the multiple criterion (trial 1, 2, and 3) and predictor variables (THEA and THEA Number), a multiple regression or canonical correlational analysis could be applied to address the research questions concerning prediction and explanation (Stevens, 1996; Thompson, 2000, 1986).

Implications of Research

When we know student misconceptions, we can design our curriculum, our assessment items, and our teaching strategies to elicit these ideas and challenge them (Cobb, 2000; Minstrell, 2002, 2001). Elaboration activities offer an opportunity to test the reliability and validity of new ideas and to explore contexts of application of the new ideas. Assessment embedded in instruction allows students to check on their understanding and allows teachers to monitor progress and identify instructional needs of individuals or of the whole class (Fuchs & Fuchs, 1986; Kulm, 1996; Minstrell, 2001;

Standards, 2000). In addition to benefits of analysis of student cognition, this research offers empirical data that supports developmental research (Cobb, 2000).

References

Caverly, D., & MacDonald, L. (Fall2003). How technology has changed developmental education. *Journal of Developmental Education*, 27, 1, p38-41.

Clements, D. & Battista, M. (2000). Designing effective software. In A. E. Kelly & R.A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 761-776). London: Erlbaum.

Cobb, P. (2000). Conducting teaching experiments in collaboration with teachers. In A. E. Kelly & R.A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 307-333). London: Erlbaum.

Fuchs, L., & Fuchs, D. (1986). Effects of systematic formative evaluation: A Meta-Analysis. *Exceptional Children, 53*, 199-208.

Guttman, L. (1958). What lies ahead for factor analysis? *Educational and Psychological Measurement, 18*, 497-515.

Kulm, G. (1994). *Mathematics assessment: What works in the classroom.* San Francisco: Jossey-Bass.

Levy, S. (ed.) (1994). *Louis Guttman on theory and methodology: Selected writings*. Aldershot: Dartmouth.

MacDonald, J. (2001). Exploiting online interactivity to enhance assignment development and feedback in distance education. *Open Learning, 16*, 179-189.

MacDonald, L., Vasquez, S., & Caverly, D. (Winter2002). Effective technology use in developmental mathematics. *Journal of Developmental Education*, *26*, 2, 36-40.

Minstrell, J. (2000). Student thinking and related assessment: Creating a facet assessment-based learning environment. In J. Pellegrino, L. Jones, & K. Mitchell (Eds.), *Grading the nation's report card: Research from the evaluation of NAEP.* Washington D.C.: National Academy Press.

Minstrell, J. (2001). The role of the teacher in making sense of classroom experiences and effecting better learning. In D. Klahr & S. Carver (Eds.), *Cognition and instruction: 25 years of progress.* Mahwah, NJ:

Minstrell, J. (2002). Facets of student's thinking: Designing to cross the gap from research to standards-based practice. In K. Crowley, C. Schunn, & T. Okada (Eds.), *Designing for science: Implications from professional, instructional, and everyday science.* Mawah, NJ: Lawrence Erlbaum Associates.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Research Council. (1999). How *people learn: Brain, mind, experience, and school*. Washington DC: National Academy Press.

Research Advisory Committee. (1998). New challenges to the research community: Reflections of the research advisory committee, *Journal for Research in Mathematics Education, 29*, 201-227.

Romberg, T. (1992). *Mathematics assessment and evaluation: Imperatives for mathematics educators*. Albany: State University of New York.

Schumacker, R. (1999). Many-facet Rasch analysis with cross, nested, and mixed designs. *Journal of Outcome Measurement, 3*, 323-338.

Shye, S. (ed.) (1978). *Theory construction and data analysis in the behavioral sciences*. San Francisco: Jossey-Bass.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Hillsdale, NJ: Erlbaum.

Thompson, B. (2000). Canonical correlation analysis. In L. Grimm & P. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 285-316). Washington, DC: American Psychological Association.

Thompson, B. (1984). *Canonical correlation analysis: Uses and interpretation*. Newbury Park, CA: SAGE.

Zvulun, E. (1978). Multidimensional scalogram analysis: The method and its application. In Shye, S. (ed.) (pp. 237-255). *Theory construction and data analysis in the behavioral sciences*. San Francisco: Jossey-Bass.