# An Alternative Mathematics Assessment Mode: Mathematical Modeling

Ozgul Kartal

Illinois Institute of Technology

Beyza Dunya

University of Illinois at Chicago

**Abstract**

This study investigates whether performance on an alternative mode of assessment may provide a means to identify mathematical capabilities that are not captured on conventional measures of mathematics achievement. Mathematical modeling is proposed as an alternative mode of assessment that has potential to become part of a suite of assessment modes that could contribute to a rich profile that can better identify students who have the capability to pursue STEM career paths. Findings of the study provide compelling evidence for the argument that mathematical modeling activities assess capabilities different than those tapped by traditional assessment, and thus can be considered a powerful alternative form of assessment.

An Alternative Mathematics Assessment Mode: Mathematical Modeling

## Purpose of the Study

Disproportionate numbers of disadvantaged students who perform below norms on traditional tests of mathematical competency drop out of mathematics and are thus denied access to important skills and pathways to economic and other types of enfranchisements (Madison & Hart, 1990; Miller, 1995; National Action Committee for Minorities in Engineering, 1997; National Commission on Mathematics and Science Teaching for the 21st Century, 2000; National Science Foundation [NSF], 2000). The worrisome fact is that some of those who encounter and fail to perform at high levels on traditional high-stakes tests may indeed be capable of success in science, technology, engineering, and mathematics (STEM) fields of study and ultimately STEM careers. The research of Carraher, Carraher, and Schliemann (1985) exemplifies the problem. They identified students who demonstrated high levels of mathematical proficiency in real-life contexts, yet found those same students while in school did not perform as well on context-free paper-and-pencil tests of corresponding skills.

The purpose of this study is to systematically investigate whether performance on an alternative mode of assessment may provide a means to identify mathematical capabilities that are not captured on conventional measures of mathematics achievement. For example, National Council of Teachers of Mathematics (NCTM, 2012) states that ongoing formative and summative assessments should provide students with opportunities to demonstrate mathematical knowledge in multiple ways.  The council recommends that decisions about how well students are performing should be made on the basis of a variety of assessments. Moreover, NCTM's (2013) position statement on formative assessment is that formative strategies should provide opportunities for students to make conjectures, incorporate multiple representations in their problem solving, and discuss their mathematical thinking with their peers. In addition, given that current national goals in the United States (e.g., National Science Foundation [NSF], 2013) include an emphasis on increasing diversity in the population of STEM professionals, finding ways to broaden the suite of assessments that are typically used for gate keeping to such professions is a critical piece of the puzzle.

## Theoretical Framework and Literature Review

Researchers have long been concerned that traditional assessment modes and instruments that are predominant in mathematics education fail to provide valid insight into the full range of what students know, understand, and can achieve, in particular as far as higher order thinking, insight, and ability are concerned (e.g., Leder, Brew, & Rowley, 1999; Niss, 1999; Lesh & Sriraman, 2005; Simon & Forgette-Giroux, 2000; Stephens, 1987; Watt, 2005). Traditional assessment modes and instruments—as described by Schoenfeld (2002)— are standardized tests comprised of selected or short-response items, or teacher made tests in which students typically need to perform some computations and arrive at the one single correct answer to earn full credit.

Two unintentional consequences of traditional assessments are discussed by researchers and educators.

First, by pointing to the differences in learner characteristics, researchers state that over-reliance on any one form of assessment (e.g., traditional assessment) disfranchises students who are able to display their knowledge, skills, or abilities more effectively through other methods (e.g., Leder et al., 1999; Niss, 1999; Stephens, 1987; Schoenfeld, 2002). As argued by Schoenfeld (2002), failing students in mathematics based on this one form of assessment closes off an important means of access to society’s resources. In particular, society loses the opportunity to benefit from a diversity of perspectives in fields heavily reliant on mathematical thinking such as engineering (Frehill, Di Fabio, & Hill, 2008). As a matter of fact, numerous professional organizations, including American Educational Research Association (AERA, 2000) and NCTM(2000), consistently recommend that alternative assessments should be used to provide complete and accurate reflections of students' abilities; that assessments should cover the broad spectrum of content and thought processes represented in the curriculum, not simply those that are easily measured; and that tests must provide appropriate accommodations for all students with different learning characteristics. Thus, there needs to be a wide range of methods for gathering assessment information that encompass a range of learning styles and capabilities.

A second unintentional consequence is that intended mathematics educational objectives may not be achieved when performance is evaluated by only traditional assessments (e.g., Lesh & Clarke, 2000; Stephens, 1987). More specifically, Stephens describes how some instructional goals are likely to be emphasized and others de-emphasized due to the reliance on one form of assessment, rather than reliance on a variety of assessment modes that can capture the breadth and depth of a curriculum. Many researchers agree that traditional mathematics assessments typically focus on low-level facts, repetition of learned procedures, and routine skills and algorithms using small sets of problems (e.g., Lesh & Clarke, 2000; Clarke & Lovitt, 1987; Grimison, 1992; Firestone, Winter, & Fitz, 2000). A mathematics education experience that tightly coordinates implemented instructional goals with these types of tests leads to an impoverished curriculum.

This study focuses on a type of alternative assessment that has potential to become part of a suite of assessment modes that could contribute to a rich profile of students who have the desire to pursue STEM career paths. The focus is on student creation of mathematical models in response to complex interdisciplinary problems. One reason for this focus is that "modeling" is one of the eight mathematical practices that are emphasized in the CCSSM across all grades—consistent with the importance of this aspect of mathematical practice in future-oriented curricula. Requiring students to create or adapt mathematical models to solve complex problems parallels the real world work encountered in many STEM careers, such as applied mathematics (Lesh & Doerr, 2003) and engineering (Gainsburg, 2007). Yet, in the literature there are frequent references to discrepancies between students' performance on realistic modeling problems compared to their performance on traditional assessment. Many of the claims in the literature are anecdotal and case-based (e.g., Lesh & Harel, 2003; Lesh & Sriraman, 2005), although one study by Iversen and Larson (2006) provides a direct comparison with a large sample. In particular, they report that different capabilities are tapped by conventional tests compared to assessments based on creating mathematical models for interdisciplinary problems.

Given that there appears to be discrepancies between students' performance on traditional assessments and mathematical modeling of complex problem settings, that there is inadequate evidence about nontraditional students modeling performance, that modeling is recognized as an important component of mathematical capability in future-oriented STEM careers, and that nontraditional populations of students (i.e., women) find mathematical modeling activity to enhance interest in STEM coursework, the following research question drives the design of this study:

> What is the relationship between students' performance on traditional assessments in mathematics and the quality and nature of the mathematical models they create for realistic problem situations?

# Methods

Students' SAT scores are used as indicators of traditional mathematics achievement. The models students created for model-eliciting activities (MEAs) were used in this study to investigate the potential for assessment of modeling capabilities as a viable alternative mode of assessment to traditional tests.

## Sample

The data for this study were collected during spring 2012 semester from a total of 1655 students enrolled in an undergraduate course. The course met in sections of 120 students (maximum) twice each week for 110-minutes. Instruction was faculty led and supported by one Graduate Teaching Assistant (GTA) and four Peer Teachers (PTs) ranging from sophomores to fifth-year seniors. Two MEAs were implemented during the semester-long course; each was launched in the classroom and iteratively modified to completion using peer and instructor feedback outside of class by students working in teams of three or four. Thus, there were a total of 416 teams across 15 sections.

## MEAs

MEAs are interdisciplinary realistic problems in which a client expresses a need for a solution to a complex problem that requires a mathematical model be produced. MEAs are carefully designed based on six design principles (Lesh, Hoover, Hole, Kelly, and Post, 2000) and repeatedly field tested until they do indeed prompt students to generate mathematical models when students are genuinely engaged in the problems.

A typical format of an MEA, as described by Diefes-Dux, Hjalmarson, Miller, and Lesh (2008), is that the students first read an article or a description that helps them enter into the MEA problem context. This is followed by the MEA problem statement, a memo from the client expressing the need for a mathematical model. The MEA problem statement is written in a way that requires the students define for themselves the problem that the client needs solved. Then students collaborate with peers to create a mathematical model that will successfully meet the client's needs. During this collaborative process, problem solvers typically describe, revise, and refine their ideas during the problem-solving episode and use of a variety representational media to explain (and document) the conceptual systems they have designed (Lesh, Carmona, & Post, 2004). Typically, one episode lasts a couple of weeks. A variety of reasonable models can be designed to meet the client's needs when assumptions and rationales are well articulated. Students who productively engage in the MEA typically go through multiple iterations of testing and revising their solution (i.e., models), ensuring that their procedure will be useful to the client.

The MEA used in this study was Just-In-Time Manufacturing. The task was to develop a procedure to rank potential shipping companies using historical data. The historical data provided to students were the numbers of minutes late the potential companies' deliveries arrive. There were eight shipping companies and 255 data points for each of the shipping companies. The students were also asked to demonstrate the functionality of their procedure and to include their reasoning for their procedure.

**Data**

Two sets of data were used to answer the research questions. The first set of data came from the students' SAT scores. This first set of data was used as an indicator of students' achievement as measured by traditional assessments. The second set of data is students' scores on the MEA. The MEA was scored using the four-dimension rubric comprised of seven components, each with a maximum score of four. The rubric and evaluation method has been described in detail by Verleger et al., (2010). The first dimension was mathematical model with two components that assess how well the mathematical model addresses the complexity of the problem and how well the procedure takes into account all types of data provided in the problem. The second dimension, reusability, is assessed with one component that looks at how well the problem is articulated. The third dimension, shareability, has three components that assess how well the results are presented, and the ease with which the model can be used to reproduce the results, and the lack of extraneous information. The last dimension, modifiability, is assessed with one component that looks at how well the critical steps in the procedure are supported with rationales. The minimum of the seven component scores were taken as the combined score for the MEA as the model is only as good as its weakest element.

## Analysis

We employed the statistical technique of ordinal logistic regression to determine whether the traditional indicators of the math achievement can predict students' and teams' MEA performances. The dependent variable, MEA scores of the students, was on an ordinal scale in which the lowest value defines the lowest achievement level. The individual SAT scores and the highest and lowest SAT scores within the teams as predictors of the team's MEA score were added to the regression model as continuous predictor variables.

Then, using chi-square contingency table analysis, MEA score frequencies were compared between the low and high SAT groups.

## Findings

The first ordinal logistic regression analysis was conducted to investigate if students' individual math achievement as measured by the SAT predicted their performance on the modeling problem—MEA—as the outcome variable. The critical assumption of the ordinal logistic regression, parallel lines, was held by $\chi^2(3, N = 1251) = 4.581, p = .205$. This non-significant test of parallel lines assured that using ordinal regression is appropriate for the particular sample (Cohen et al., 2003).

The ordinal logistic regression results showed that the SAT was not a significant predictor of MEA performance of the individual students, $\chi^2(1, N= 1251)= 1.509, p= .219$. The ordinal logistic regression was then conducted at the team level in consideration of within group differences in SAT scores. Thus, the highest and lowest SAT scores of each team were used as two predictors and another ordinal regression analysis was conducted to see whether they predicted the teams' MEA-1 scores. The assumption of the parallel lines, was held by $\chi$ 2$(6, N = 174) =3.926, p = .687$. The non-significant test of parallel lines assured that using ordinal regression is appropriate for this particular sample. The main results indicated that both the highest and the lowest SAT scores were not significant predictors of the teams' MEA-1 performance. The overall regression model was non-significant $\chi^2(2, N=174) = .561, p= .755$.

A chi-square analysis was conducted to investigate whether the low- and high-achievers on the SAT performed differently on MEA-1. The chi-square contingency table did not yield any significant difference between low and high SAT performers' scores on MEA-1, $\chi^2(8, N = 1251) = 6.35, p = .6$.

The majority (75.4%) of the SAT low-achiever group performed at the level of 60 and 70, where only a nonsignificant fraction of the group (0.7%) performed at the level of zero, and 21.6% performed low (at the level of 50) on the MEA-1. Only 6.5% of the students at the level of 0 on MEA-1 were SAT low-achievers, and the majority (93.5%) of the level 0 performers was comprised of SAT high- or medium-achievers. Findings showed that 22% of the students at the level of 50 on MEA-1 were SAT low-achievers.

Representation of SAT high-achievers' performance on MEA-1 is not different than the SAT low-achievers group. The majority (71.9%) of the group performed at the level of 60 and 70, where only a nonsignificant fraction of the group (0.3%) performed at the level of zero, and 5.5% performed low (at the level of 50) on the MEA-1. The data showed that 3.2% of the students at the level of 0, and 6% of the students at the level of 50 on MEA-1 were SAT high-achievers.

## Conclusion

One major finding of this study is that students' math achievement as measured by a traditional assessment, the SAT, did not predict their performance on an alternate form of assessment, the MEA. This finding lends support to the belief that traditional assessment modes and instruments that are predominant in mathematics education fail to provide a complete picture of what students know, understand, and can achieve, especially with respect to important capabilities for success in STEM professions (e.g., modeling).

Another major finding is that there is no difference between low and high traditional-test achievers' performance on the MEA. More importantly, the majority of the low traditional-test achievement group performed high on the MEA. This finding provides evidence to support Lesh and Sriraman's (2005) claim that low performance on traditional tests do not always coincide with low mathematical problem-solving and modeling abilities, and that traditional assessments fail to identify students who can powerfully and effectively apply mathematics to real-world problems, such as MEAs.

This study provides compelling evidence to the argument that MEAs assess capabilities different than those tapped by traditional assessment, and thus can be considered a powerful alternative form of assessment. MEAs can complement traditional forms of assessment in order to reflect a more complete and accurate picture of students' mathematical abilities, to address a range of learning styles, and to possibly open pathways for nontraditional populations into the studies for STEM careers. Further, the results suggest a viable alternative assessment mode for identifying a broader diversity of talented students who might otherwise be denied access to STEM oriented tracks of study.

## References

American Educational Research Association. (2000). *Ethical standards of the American Educational Research Association.* Retrieved July 28, 2007, from http://www.aera.net/uploadedFiles/About_AERA/Ethical_Standards/EthicalStandards.pdf

Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (1985). Mathematics in the streets and in schools. *British Journal of Developmental Psychology*, *3*(1), 21-29.

Clarke, D., & Lovitt, C. (1987). MCTP assessment alternatives in mathematics. *Australian Mathematics Teacher 43*(3), 11–12.

Common Core State Standards Initiative. (2010). *Common Core State Standards for mathematics*. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Cohen, J., Cohen, P., West, G. S., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* New York, NY: Routledge.

Diefes-Dux. H. A, Hjalmarson, M. A., Miller, T. K., & Lesh, R. (2008). Chapter 2: Model-Eliciting Activities for engineering education. In J. S. Zawojewski, H. A. Diefes-Dux, & K. J. Bowman (Eds.), *Models and modeling in Engineering Education: Designing experiences for all students* (pp. 17-35). Rotterdam, the Netherlands: Sense Publishers.

Firestone, W.A., Winter, J., & Fitz, J. (2000). Different assessments, common practice? Mathematics testing and teaching in the USA and England and Wales. *Assessment in Education,* 7(1), 13–37.

Frehill, L. M., Di Fabio, N. M., & Hill, S. T. (2008). *Confronting the new American dilemma: underrepresented minorities in engineering: A data-based look at diversity*. Washington, DC: National Action Council for Minorities in Engineering.

Gainsburg, J. (2007). The mathematical disposition of structural engineers. *Journal for Research in Mathematics Education*, *38*(5), 477-506.

Grimison, L. (1992). *Assessment in mathematics: Some alternatives*. Paper presented at the Mathematics Education Research Group of Australia 15th Annual Conference, Sydney, Australia.

Leder, G. C., Brew, C., & Rowley, G. (1999). Gender differences in mathematics achievement: Here today and gone tomorrow? In G. Kaiser,E. Luna & I. Huntley (Eds.), *International comparisons in mathematics education* (pp.213-224). London, UK: Falmer Press.

Lesh, R., Carmona, G., & Post, T. (2002) *Models and modeling*. Paper presented at the Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Athens, GA.

Lesh, R., & Clarke, D. (2000). Formulating operational definitions of desired outcomes of instruction in mathematics and science education. In Kelly, A. E. & Lesh, R. (Eds.). *Handbook of Research Design in Mathematics and Science Education* (pp. 113-151). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lesh, R., & Doerr, H. M. (2003). Foundations of a models and modeling perspective on mathematics teaching, learning, and problem solving. In R. Lesh & H. M. Doerr (Eds.), *Beyond Constructivism: Models And Modeling Perspectives On Mathematics Teaching, Learning, And Problem Solving* (pp. 3-33). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lesh, R., & Harel, G. (2003). Problem solving, modeling, and local conceptual development. *Mathematical Thinking and Learning*, *5*(2-3), 157-189.

Lesh, R., Hoover, M., Hole, B., Kelly, A., & Post, T., (2000). Principles for developing thought-revealing activities for students and teachers. In A. Kelly, & R. Lesh (Eds.), *Research Design in Mathematics and Science Education* (pp. 591-646). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lesh, R., & Sriraman, B. (2005). Mathematics education as a design science. *ZDM*, *37*(6), 490 505.

Madison, B. L., & Hart, T. A. (1990). *A challenge of numbers: People in the mathematical sciences.* Washington, DC: National Academy Press.

Miller, S. (1995). *An American imperative.* New Haven, CT: Yale University Press.

National Action Committee for Minorities in Engineering. (1997). *Engineering and affirmative action: Crisis in the making.* New York: Author.

National Commission on Mathematics and Science Teaching for the 21st Century. (2000). *Before it's too late: A report to the nation.* Washington, DC: U.S. Department of Education.

National Council of Teachers of Mathematics (2012). *Large-Scale Mathematics Assessments and High-Stakes Decisions*. Reston, VA: Author.

Niss, M. (1999). Aspects of the nature and state of research in mathematics education. *Educational Studies in Mathematics*, *40*(1), 1-24.

Schoenfeld, A. H. (2002). Making mathematics work for all children: Issues of standards, testing, and equity. *Educational researcher*, *31*(1), 13-25.

Simon, M., & Forgette-Giroux, R. (2000). Impact of a content selection framework on portfolio assessment at the classroom level. *Assessment in Education: Principles, Policy & Practice*, *7*(1), 83-100.

Stephens, M. (1987). Towards an AAMT policy on assessment and reporting in school mathematics. *Australian Mathematics Teacher*, *43*(3), 2–3.

Verleger, M., Diefes-Dux, H., Ohland, M. W., Besterfield-Sacre, M., & Brophy, S. (2010). Challenges to informed peer review matching algorithms. *Journal of Engineering Education*, *99*(4), 397–408.

Watt, H. M. (2005). Attitudes to the use of alternative assessment methods in mathematics: A study with secondary mathematics teachers in Sydney, Australia. *Educational Studies in Mathematics*, *58*(1), 21-44.