



International Journal of Research in Education and Science (IJRES)

The Relationship between Students' Performance on Conventional Standardized Mathematics Assessments and Complex Mathematical Modeling Problems

Ozgul Kartal¹, Beyza Aksu Dunya², Heidi A. Diefes-Dux³
Judith S. Zawojewski⁴

¹Saint Xavier University, ozgulkartal@gmail.com

²University of Illinois at Chicago, baksu2@uic.edu

³Purdue University, hdiefes@purdue.edu

⁴Illinois Institute of Technology, judijz@gmail.com

www.ijres.net

To cite this article:

Kartal, O., Dunya, A.B., Diefes-Dux, A.H., & Zawojewski, S.J. (2016). The relationship between students' performance on conventional standardized mathematics assessments and complex mathematical modeling problems. *International Journal of Research in Education and Science (IJRES)*, 2(1), 239-252.

This article may be used for research, teaching, and private study purposes.

Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

Authors alone are responsible for the contents of their articles. The journal owns the copyright of the articles.

The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of the research material.

The Relationship between Students' Performance on Conventional Standardized Mathematics Assessments and Complex Mathematical Modeling Problems

Ozgul Kartal^{1*}, Beyza Aksu Dunya², Heidi A. Diefes-Dux³, Judith S. Zawojewski⁴

¹Saint Xavier University, ²University of Illinois at Chicago, ³Purdue University, ⁴Illinois Institute of Technology

Abstract

Critical to many science, technology, engineering, and mathematics (STEM) career paths is mathematical modeling—specifically, the creation and adaptation of mathematical models to solve problems in complex settings. Conventional standardized measures of mathematics achievement are not structured to directly assess this type of mathematical modeling. Therefore, a major question is whether a conventional standardized test can serve as a reliable *predictor* of students' potential to mathematical modeling performance. To investigate this question, a study was designed to find the relation between students' conventional standardized measures of mathematics achievement and their performance on mathematical modeling problems. Students' (N=1656) SAT (Scholastic Aptitude Test) mathematics scores were used as a conventional standardized measure of achievement and students' scores on two model-creation problems based on complex settings were used to capture mathematical modeling performance to answer the question whether standardized achievement tests function well in predicting their mathematical modeling performance.

Key words: Mathematical modeling; Model eliciting activities; STEM careers; Modeling capabilities, Standardized tests

Introduction

Mathematical modeling is currently important and gaining importance in STEM fields during this time of rapidly changing and emerging fields. For example, nanotechnology, which is the study of how to manipulate matter at the molecular and atomic levels, uses mathematical modeling to help theoretically understand the behavior of nanomaterials and reduce the costs of building numerous prototypes. As such, developing students' mathematical modeling abilities in context is critical to engineering career paths, as reflected in the criteria set by accreditation body (ABET) (2010) for engineering education programs and in the vision of education from the Engineer of 2020 by the National Academy of Engineering (NAE) (2005). Similarly, modeling of complex situations is increasingly being used in other scientific fields, such as in molecular cell biology, where mathematical models serve as working hypotheses; they help understand and predict the behavior of complex systems. Therefore, a major goal, for many, if not most, STEM career paths is to develop professionals who can engage successfully in designing mathematical models for complex situations.

Although past and current national goals in the United States (e.g., National Science Foundation [NSF], 2013) include an emphasis on increasing diversity in the STEM professions, conventional standardized mathematics test, often used as important gatekeepers to the fields, can act as barriers to this very goal. Indeed, a disproportionate number 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 Board, 2000). The worrisome fact is that some of those who fail to perform at high levels on traditional high-stakes tests may indeed be capable of success in STEM fields of study and careers. The research of Carraher, Carraher, and Schliemann (1985) exemplifies the problem. They identified students who demonstrated high levels of mathematical proficiency in verbal contexts associated with real life problem solving, yet found those same students while in school did not perform as well on context-free paper-and-pencil tests on corresponding skills. Further, some anecdotal evidence (e.g., Lesh & Harel, 2003;

* Corresponding Author: Ozgul Kartal,, ozgulkartal@gmail.com

Lesh & Sriraman, 2005) and empirical evidence (Iversen and Larson, 2006) suggests that when asked to engage in mathematically modeling complex situations, unexpected students emerge as talented—not necessarily those who perform well on conventional standardized mathematics assessments.

The dilemma is that while conventional standardized tests provide a reasonably quick and inexpensive way to sort students into appropriate programs of study, such tests are not structured to directly assess students' performance on complex mathematical activity, such as modeling. One problem is that the time needed to create or adapt a mathematical model to a complex problem situation may take 15 to 90 minutes, which is inconsistent with the need to administer a relatively large number of independent items to achieve psychometric reliability. Another structural problem is that a number of reasonable models may be devised as a solution to a complex problem, depending on the assumptions and rationales made during the design process—which introduces a challenge to the scoring of the student responses for the purpose of psychometric analysis. However, the overall concern about the mismatch between the nature of standardized tests and students' future performance on complex mathematical activity, such as modeling, could be eliminated if conventional standardized tests of mathematics achievement serve as reliable *predictors* of students' potential mathematical modeling performance. To investigate this question, this study was designed to find the relation between students' performance on conventional standardized measures of mathematics achievement and their subsequent performance on mathematical modeling problems.

The purpose of this study is to systematically investigate whether a conventional standardized test can serve as a reliable *predictor* of students' potential to mathematical modeling performance.

Theoretical Framework and Literature Review

Consequences of Reliance on Conventional Standardized Mathematics Assessments

A number of researchers have long been concerned that standardized 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; Lesh & Sriraman, 2005; Niss, 1999; Simon & Forgette-Giroux, 2000; Stephens, 1987; Watt, 2005). Many (e.g., Leder et al., 1999; Niss, 1999; Stephens, 1987; Schoenfeld, 2002) emphasize that over-reliance on any one form of assessment disfranchises students who are able to display their knowledge, skills, or abilities more effectively through other methods. 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, Frehill, Di Fabio, and Hill (2008) describe how society loses the opportunity to benefit from a diversity of perspectives in fields heavily reliant on mathematical thinking such as engineering.

Further, 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., Clarke & Lovitt, 1987; Firestone, Winter, & Fitz, 2000; Grimison, 1992; Lesh & Clarke, 2000). Others (e.g., Lesh & Clarke, 2000; Stephens, 1987) describe the consequential curricular emphasis on some goals (typically low level) and de-emphasis on others (high level) due to the reliance on these types of assessments. More specifically, a mathematics education experience that tightly coordinates implemented instructional goals with these types of tests leads to an impoverished curriculum.

An effort to support a richer mathematics pre-collegiate curriculum, as described in National Council of Teachers of Mathematics' (NCTM) 1989 standards and recommendations for school mathematics, was enacted by the NSF during the 1990s when they funded the development of NCTM-based school mathematics curricula (Schoenfeld, 2002; Senk & Thompson, 2003). At the time, NCTM (1989) endorsed "recognition of mathematics as more than a collection of concepts and skills to be mastered" and included "methods of investigating and reasoning, means of communication, and notions of context" (p.5). The more recent NCTM (2000) standards document expands on the original initiative by viewing mathematical sophistication as a core component of intelligent decision making in everyday life, in the workplace, and in our democratic society. The 2009 Common Core State Standards for Mathematics (CCSSM), recently introduced across most of the United States (U.S.), also emphasizes conceptual understanding and reasoning, career readiness, and full participation in society (NCTM, 2013). Schoenfeld (2002) and Watt (2005) situate these calls for an enriched curriculum in the 21st century, which they portray as increasingly technological in the workplace and personal life. They point to the rapidly evolving societal demands for individuals who are capable of problem solving, reasoning, making mathematical connections, communication, and working collaboratively in mathematical applications. The

following CCSSM's description of the mathematically proficient student is in alignment with Schoenfeld and Watt's discussion.

Mathematically proficient students can apply the mathematics they know to solve problems arising in everyday life, society, and the workplace. ...can apply what they know...are comfortable making assumptions and approximations to simplify a complicated situation realizing that these may need revision later. ...routinely interpret their mathematical results in the context of the situation and reflect on whether the results make sense, possibly improving the model if it has not served its purpose. (Common Core State Standards Initiative [CCSSI], 2010, p. 7)

The U.S. Department of Education has funded the development of two testing systems to align with the new curriculum standards (Cizek, 2010): The Partnership for Assessment of Readiness for College and Careers (PARCC; www.parcconline.org) and the Smarter Balanced Assessment (<http://www.smarterbalanced.org>). A question is whether any conventional standardized assessment of mathematics achievement can serve as reasonable predictors of student potential success in goals reflecting 21st Century needs, such as performance on complex mathematical modeling problems. Ridgway, Zawojewski, and Hoover (2002) discuss two challenges for developing new forms of assessment are developed: (1) there may be no way of assessing new goals using existing techniques; and (2) the relationship between new and old measures of performance is unknown. Conclusions about whether these new tests meet the challenges of assessing higher-ordered performance will not be possible until the PARCC and the Smarter Balanced Assessment have been in place for a few years and follow a period of enactment of the new CCSSM curriculum. Further, college placement tests, such as the SAT and the American College Testing (ACT) will continue to be used to influence the various career paths that students may pursue.

Mathematical Modeling and Standardized Tests

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 using 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.

The Iversen and Larson study was conducted at a university in Denmark during a seven-week calculus course enrolling about 200 students. At the beginning of the course, the students took a pre-test consisting of eleven traditional assessment problems. After the pre-test, the students were presented a modeling problem. At the end of the calculus course, the students completed the post-test—a two-hour-written-exam—consisting of a number of basic calculus problems that were similar to the problems they worked on during the course, and similar in nature to the traditional items that comprised the pre-test. Correlating students' modeling scores with pre-test scores and modeling scores with post-test scores, the researchers found that neither the post-test nor the pre-test did a good job of predicting which students were able to do well on modeling problems. Hence, Iversen and Larson concluded that the students who perform well in traditional testing environments do not necessarily succeed on more complex real-life modeling situations. In this regard, the findings of the study support Lesh and Sriraman's (2005) statement that traditional mathematics assessments often fail to identify students who can powerfully and effectively apply mathematics to real-world problems, and many students who excel on traditional assessments often struggle to implement their mathematical knowledge in real-world settings.

However, one of the major limitations of the Iversen and Larson study was that the college students were exposed to just a single modeling activity for the first and only time for the purpose of the study. This small exposure to modeling experiences may explain why students who performed well on the more traditional pre- and post-tests did not perform well on the modeling problem. In other words, their results might be different if the students were given more opportunities to solve a number of modeling problems of the type administered after the pre-test. Another limitation is that Iversen and Larson did not address how students who performed poorly on the traditional tests did on the modeling activities, which would inform the significant question raised above concerning the role of standardized testing in predicting who is most likely to perform at high levels on complex mathematical activity, such as modeling.

Diefes-Dux et al. (2004), on the other hand, directly addressed the issue of providing opportunities for nontraditional populations to maintain interest and persistence in engineering via modeling activities in their coursework. Their research questions were: "What more needs to be done to improve female students' interest and persistence in engineering?" and "How can we better understand the reasons that female students leave engineering despite good academic performance?" They described the use of four modeling problems during a first-year engineering course, making the case that contextually-driven mathematical modeling activities solved in small groups serve as a potentially productive way of addressing and assessing gender equity in the engineering classroom. The Purdue University course was required of the large number of incoming engineering students per year (approximately 1700) and addressed engineering problem solving and computer tools skills development. The findings of the study revealed that by the time students had completed four modeling activities, both men and women in the course reported positive feelings toward the activities, and that women were even more positive about the nature of the work compared to the men. Thus, modeling activities appear to have potential for enhancing at least the interest of nontraditional populations—in this case women. Furthermore, another study by Diefes-Dux, Hjalmarson, Zawojewski, and Bowman (2006) in this same project reported that engineering instructors were able to integrate complex interdisciplinary mathematical modeling activities into their coursework without compromising the required engineering content within the course.

The Research Question

The inquiry that drives this study is whether a conventional standardized test can serve as a reliable *predictor* of students' potential to mathematical modeling performance. To investigate this question, the study is designed to find the relation between students' conventional standardized measure of mathematics achievement and their subsequent performance on mathematical modeling problems. Students' SAT mathematics scores are used as a conventional standardized measure of mathematics achievement and students' scores on two model-creation problems based on complex settings within a series of four of such problems are used to represent their mathematical modeling performance. High and low scoring groups are identified from the standardized-test achievement data from a population of 1656 first-year engineering students. Their SAT scores are used as a predictor of their scores on mathematical modeling tasks to answer the question whether the standardized achievement tests function well in predicting their mathematical modeling performance. The research question is: What is the relationship between students' performance on conventional standardized assessments in mathematics and performance on modeling tasks in which they create mathematical models for realistic problem situations?

Method

To investigate the relationship between students' performance on traditional indicators of achievement and their mathematical modeling capabilities, this study taps the extensive information available concerning Purdue University first-year engineering students. Their American College Testing (ACT), Scholastic Aptitude Test (SAT), and Assessment and Learning in Knowledge Spaces (ALEKS) scores are used as indicators of traditional mathematics achievement. The two-semester long first-year course sequence from which the data was gathered incorporated four model-creation activities based on complex problem settings; these are called model-eliciting activities (MEAs). The models students created were evaluated and contributed to the course grade. Students' MEA grades are used in this study to investigate the potential for assessment of modeling capabilities as a viable alternative mode of assessment that would complement the conventional standardized tests.

Setting and Sample Population

The data was gathered from student performance in the first year engineering course, ENGR 132-Ideas to Innovations II, at Purdue University. This course is a post-requisite to ENGR 131-Ideas to Innovations I. ENGR 131 introduces students to the engineering professions using multidisciplinary, societally relevant content through developing engineering approaches to generating and exploring creative ideas, and using quantitative methods to support design decisions. In ENGR 132, students take a more in-depth approach to constructing innovative engineering solutions to open-ended problems. In each course, the students were exposed to two MEAs, addressing the lack of multiple exposures to modeling concern raised by the Iversen and Larson (2006) study. The data were collected during the spring 2012 semester from a total of 1655 students enrolled in the 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 student-generated models were 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.

Instrumentation for Assessing Modeling Capabilities

Rationale for MEAs to Assess Modeling Capabilities

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. Moore and Diefes-Dux (2004) point out that MEAs are carefully crafted to make sure that the students are given enough information to make informed decisions about when their model meets a client's stated requirements. MEAs are carefully designed based on six design principles (Lesh, Hoover, Hole, Kelly, & 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 given well-articulated varied assumptions and rationales. 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 literature reveals that the type of capabilities needed for mathematical modeling situations that are distinct from the type of skills, concepts, and procedures assessed in traditional tests. For example, Lesh, Carmona, and Post (2002) describe how modeling problems in complex interdisciplinary contexts involves higher-ordered mathematical practices such as quantifying, dimensioning, coordinating, categorizing, algebraizing, and systematizing relevant objects, relationships, actions, patterns, and regularities—all of which seldom appears on conventional standardized tests. English (2007) reported on fifth grade students' mathematical thinking and learning as they solved complex problems by creating mathematical models in small groups. In particular, she focused on how students constructed mathematical ideas and mathematized situations. English's work suggests that modeling activities reveal real-world complex problem solving processes that go beyond a single mapping from givens to goals—which is foundational to traditional testing modes.

MEAs as a form of alternative assessment has been inspired by Lesh and Clarke's (2000) statement that these activities tend to focus on problem-solving situations that involve a small number of "big ideas" that involve higher order understanding and abilities, rather than a large number and breadth of small easily-testable skills. In particular, MEAs are designed to engage students in productive mathematical thinking and in developing math concepts (Lesh, 2003; Lesh & Kelly, 1997) in the context of an interdisciplinary setting and by use of multiple mathematics topics (Lesh & Doerr, 2003). MEAs provide more insights into students' mathematical thinking than traditional tests because the very models that are created serve as windows on students' mathematical ways of thinking (Lesh et al., 2000). Unlike evidence gleaned from a student's choice of one answer from a collection of distractors, or a short answer response such as "12 miles," a student-created model provides information about student-selected mathematical elements, operations, and relations. Further, Lesh and Harel (2003) point out that the model produced as an answer for an MEA is intended to be sharable with other people, reusable in other situations, and modifiable for other purposes—helping to make students' responses clear and even more revealing of their thinking. Thus, the potential is great for drawing conclusions about what students know and can do when solving complex problems.

MEAs are also distinct from traditional instruments in that they blur the boundaries between assessment and significant learning. When actively engaged in MEAs, students engage in productive and generative mathematical thinking as revealed through the math concepts they use to create models (Lesh & Doerr, 2000; Lesh & Doerr, 2003; Lesh & Kelly, 1997). Lesh and Harel (2003) point out that MEAs require students to make

symbolic descriptions of meaningful situations, while problems on traditional assessment generally emphasize computational skills or recall of procedures. Students' production of symbolic descriptions to meaningful situation are not just answers, but rather are the critical components of conceptual tools that need to be produced to solve the problem. That is, students construct local concepts during the assessment and in so doing gain conceptual understanding through generating symbolic descriptions of a given situation.

Measuring MEA Performance for This Study

One of the advantages of using MEAs as a means to evaluate students' mathematical modeling capabilities is that assessment criteria for MEAs have been already established in the work of Diefes-Dux and colleagues. Their criteria are based on the extent to which responses meet the needs of the client—aligning the authentic problem solving perspective of the MEA experience with the evaluation process. In particular, resulting mathematical models are assessed for mathematical complexity (the integrity of the created small system involving elements, operations, and relations), the explicit statement of underlying rationales and assumptions (in order to make reasonable judgments about how well the model meets the client's needs), and quality of communication (so that the client can readily apply the model to the problem situation) (Diefes-Dux, Zawojewski, & Hjalmarson, 2010).

The two MEAs which students solved in the ENGR 132 course were used in this study. The students had already solved two other MEAs in the pre-requisite course—ENGR 131—before being exposed to these. The MEAs used in this study were: (1) Just-In-Time Manufacturing and (2) Shredded Document. In MEA-1, 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. In MEA-2, the students were given two gray scale images that were digitally shredded into 8-11 pieces. The students were asked to develop a structured method (i.e. algorithm) that uses the pixel level gray scale value information to reassemble the shredded documents and meet a target level of accuracy, so that a programming team could take their algorithm and translate it to software code. Various mathematics and science standards, as well as engineering principles, were addressed by both MEA-1 and 2 in the domains of numbers and operations, algebra, measurement, data analysis and probability, problem solving, reasoning and proof, communication, connections, representation, and inquiry. A complete list of standards addressed by each MEA is available at the MEA library by University of Minnesota, by the link <https://ay12.moodle.umn.edu/course/view.php?id=8332>.

For this setting and sample population, the scoring of students' solutions to MEAs was accomplished by a team of nine GTAs and 70 PTs using a rubric with four dimensions—mathematical model, shareability, reusability, and modifiability as described by Diefes-Dux, Zawojewski, Hjalmarson, and Cardella (2012). The GTAs and PTs were trained to employ the Instructor's MEA Assessment/Evaluation Package (I-MAP) for each MEA. An I-MAP is a MEA-specific guide for applying the generic MEA rubric to student work (Diefes-Dux et al., 2010). GTAs and PTs engaged in seven hours of training that encompassed face-to-face training and pre and post training activities performed on their own time. The MEA training model involved: (1) practice with an activity like a student, (2) exposure to the research-base and/or theoretical underpinnings, (3) practice with interpreting student work, and (4) reflective comparison to an expert (Verleger & Diefes-Dux, 2013).

Indicators for Conventional Standardized Test of Mathematical Achievement

The population of students in this study came from varied states and countries, and independent school districts in the state, and therefore the data for individual students included any one to three of the following conventional standardized tests: ACT, SAT, and ALEKS. Each of these is a standardized test for college admissions and/or mathematics placement exams for college entering students in the United States. An initial step in the analysis was to select which of the three tests to use (based on whether the tests represented similar capabilities and which tests was most prevalent among the students). So, initially, it was assumed that each of these three tests could be used as an indicator for performance on a conventional standardized assessment of mathematical achievement for the current study. The standards met by the two MEAs correspond to the same college and career readiness standards that are intended to be measured by ACT/SAT/ALEKS, yet the purposes and forms of standardized assessments are quite different.

The ACT is a standardized, multiple-choice test that assesses students' academic readiness for college. ACT has been used in admissions for four-year U.S. colleges since 1959, although research has shown that the test is not

a good predictor of college success for many minority students (Myers & Pyles, 1992). Detailed information about ACT can be found at act.org. The SAT was developed and revised by the Educational Testing Service (1948-1990). In contrast to the ACT, in math section of the SAT, students are required to work out some answers entirely on their own. There are 44 standard multiple choice questions and 10 student-produced response questions on a typical SAT test. Although the usefulness of the SAT for predicting success in college has been debated, selective institutions typically ask applicants for SAT scores (Baron and Norman, 1992). ALEKS is an assessment mechanism used in higher education to suggest the placement of incoming students on a continuum between college algebra and calculus courses (Reddy and Harper, 2013). ALEKS is composed of student-produced response questions. Carpenter and Hanna (2006) have shown that ALEKS can successfully serve as a preparedness measure for placement in a calculus course. Detailed information about ALEKS can be found in Falmagne and Doignon (2011).

Statistical Methods

The general research question, “What is the relationship between students’ performance on conventional standardized assessments in mathematics and performance on modeling tasks in which they create mathematical models for realistic problem situations?,” is parsed into a series of sub-questions that can be directly linked to the statistical analysis. The sub-questions are:

i) What is the relationship between students’ standardized test scores—ACT, SAT, and ALEKS?

This question is important in answering the overall research question as it reveals if the three conventional standardized tests are measuring similar mathematical skills and capabilities.

ii) What is the relationship between students’ scores on mathematical modeling problems—MEA-1 and MEA-2?

This question reveals if the two MEAs are measuring similar mathematical skills and capabilities—specifically, modeling capability.

iii) What is the relationship between students’ conventional standardized assessment scores and their scores on mathematical modeling problems?

This question answers the overall research question.

iv) How do low-achievers and high achievers, as measured by traditional tests, perform on mathematical modeling problems?

It is important to answer question (iv) in addition to the question (iii), because there is a conjecture in the literature that MEAs allow students with different backgrounds to emerge as talented

Data

Two sets of data were used to answer the research questions. The first set of data came from the students’ ACT, SAT, and/or ALEKS scores. This set of data was used to select an indicator of students’ mathematics achievement as measured by conventional standardized tests. The second set of data is students’ scores on the two MEAs. Each 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 Diefes-Dux, Zawojewski, Hjalmarson, and Cardella (2012). A copy of the rubric is provided in Appendix A. 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 (i.e., model) 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 model is articulated—so that another person can “use” the model to solve the problem. 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 (or model) 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 component. For example, a well-written model is not useful to the client if the mathematics employed do not address the complexity inherent to the problem; alternatively, a solid math model is not useful if the client cannot follow it successfully to completion. Since four of the seven component scores are more linked to communication than math skills, and hence what is dragging the score down may or may not be math related, only the three component scores which are math related (i.e., model complexity, data types, and modifiability) were used for this study in order to more closely align what was being assessed in the mathematical modeling assessment with what was being assessed with the conventional mathematics assessments. The minimum of these three component scores were taken as the combined score for the MEA for the purpose of this study. The MEA scores were then, converted to a five-

point scale by assigning a 50 to a score of 1, 60 to a score of 2, 70 to 3, and 80 to 4. There were also cases where a team had zero. The only way to get a zero was if the team really had not progressed far enough to even have the beginnings of a math model. These zero scores remained zero in the data.

Given that solution strategies to the MEAs were not pre-taught, there is no classroom/section effect. Rather, each team of students independently experienced the MEA episode as an entity. Given that each student—coming from a large number of high schools—experienced the conventional standardized tests independent of the higher education institution, there is no classroom/section/team effect. Therefore, the unit of analysis is the team or the student at different phases of analyses.

Analysis

The analysis was designed to answer the overall research question—what is the relationship between students' performance on conventional standardized assessments in mathematics and performance on modeling tasks in which they create mathematical models for realistic problem situations?—by considering the four sub-questions listed above. The data were initially analyzed to answer the first sub-question, by conducting pair wise correlations of the three conventional standardized test scores for each student. The correlations were then used to determine whether the three standardized tests were measuring similar mathematical skills and capabilities. The data were analyzed to answer the second sub-question, by determining correlations between individuals' two MEA scores. The correlations were used to determine if the two modeling tasks were measuring similar capabilities (i.e., modeling capability). In addition, frequencies of the standardized test scores and MEA scores were obtained to present a summary of the scores of the sample. The first two analyses were then used to determine whether to conduct analyses for the 3rd and 4th sub-questions using all three standardized tests and both modeling tasks or one test and one modeling task. For the third sub-question, the relation between the standardized test scores and the MEA scores was investigated. For the last sub-question, the differences between low and high traditional-test achievement groups on the modeling tasks were explored.

Sub-question I: Selecting an Indicator for a Conventional Standardized Test of Mathematics Achievement

The data showed that some students took all three traditional math achievement tests—SAT, ACT, and ALEKS—while some took only one or two of the tests. Therefore, a decision needed to be made about which test to use as the main indicator. The first step was to determine whether or not the three tests reveal similar information about students by describing the direction and strength of the relationship between each pair of the three scores. A strong positive relationship would show that these three tests indicate similar information about the mathematics achievement level of a student, whereas a weak or a negative relationship would show that the tests indicate different information about student's mathematics achievement level. To this end, first, bivariate correlations using Pearson product-moment correlation coefficients were calculated for SAT, ACT, and ALEKS scores of each student who took all three of them. After finding that the three tests have moderate to high correlation between each other, a second step in the decision process involved examining the frequencies for each of the three conventional standardized tests to determine which test was taken by the greatest number of students. The SAT was found to be the most frequently taken test across the participants, and was therefore used as an indicator of conventional standardized test of mathematics achievement. And, any subject who did *not* have an SAT score was eliminated from the sample.

Sub-question II: Selecting an Indicator for Modeling Performance (i.e., MEAs)

Students' performance on MEA-1 and MEA-2 were correlated to find out if students' scores were consistent across the two MEAs, so as to determine whether one or both of the MEA scores would be used to indicate performance on modeling problems. The Spearman's Rho (r_s) correlation coefficient was used to determine the strength of the relation between the two MEA scores of the students. A high positive correlation would allow us to select one of the two for the analysis to avoid repetition of the results. A low or a negative correlation would show that the two MEAs each measure significantly different capabilities which would require us to analyze and report on both.

Sub-question III: Relations between Standardized Test Performance and Modeling Performance (i.e., MEA) Scores

We employed the statistical technique of ordinal logistic regression to determine whether a conventional indicator of the mathematics achievement can be used to predict students' and teams' performances on modeling problems (i.e., MEA). The rationale for using ordinal logistic regression rather than simple and/or multiple linear regression was that the MEA scores had five numerical categories; 0, 50, 60, 70, and 80. When the ordinal dependent variable (i.e., the MEA score) has five categories or less, treating this variable as continuous and running linear regression may distort the findings (Torra, Domingo-Ferrer, Mateo-Sanz, & Ng, 2006). Ordinal logistic regression allows for modeling the dependence of an ordinal outcome on one or more predictors. The dependent variable, MEA scores of the students, was on an ordinal scale in which the lowest value defines the lowest achievement level.

Another concern was that the MEA scores are at the team level whereas SAT scores are at the individual level. In an effort to overcome the difficulty of comparing individual SAT scores with team MEA scores, we considered the approach used by Iversen and Larson, wherein they analyzed their individual and team data by correlating individual MEA scores and individual traditional test scores, and correlating group MEA scores and the sum of the individuals' traditional scores in each group. However, taking the sum of team members' traditional test scores and conducting a team-level analysis gives rise to the question: Do the teams perform at the level of their mean? Perhaps, they perform at the level of their most successful or least successful member. Aggregating the SAT scores from teams with substantially low and high SAT scores may distort the results. Therefore, three different continuous predictor variables related to SAT scores were included in the regression model: the individual SAT scores (as done in Iversen and Larson's (2006) analysis); the high SAT score for each team; and the low SAT score for each team. The latter two variables were intended to detect if extremely low or high scores distorted the findings. Before performing the regression analysis, the critical assumption of ordinal logistic regression, parallel lines, was tested. This assumption indicates that the predictor variable (SAT scores) has the same impact on crossing all the categories of the dependent variable (MEA scores) (Cohen, Cohen, West, & Aiken, 2003).

Sub-question IV: Relations between Low- and High Achievers' on Traditional Tests and their Performance on Alternative Assessment (I.E., MEA)

Answering the fourth sub-question was important as it would provide evidence for the claim that students with different mathematical backgrounds emerge as talented when they are engaged in mathematical modeling. In order to group students as high/low-achievers on the SAT, the 25th and 75th percentile ranks were calculated for the SAT scores of the sample. The SAT score variable was then re-coded and transformed into a grouping variable where 1 corresponded to a score in the 25th percentile, 2 corresponded to a score in between the 25th and 75th percentile, and 3 corresponded to a score in the 75th percentile. Thus, low performers are in the first percentile group, and high performers are in the third percentile group. Descriptive statistics and frequency statistics were displayed in particular for the low and high traditional-test achievement groups in order to describe the trends among the groups.

Results

Findings for Sub-question I: Selecting an Indicator for a Conventional Standardized Test of Mathematics Achievement

Frequency tables showed that 1217 students provided their SAT scores, 798 students provided their ACT scores, and 1123 provided their ALEKS scores. A high correlation was found between students' SAT and ACT scores with $r(672) = .66, p < 0.05$, and a moderate to high correlations were detected between their SAT and ALEKS scores with $r(1018) = .50, p < 0.05$, and their ACT and ALEKS scores with $r(764) = .55, p < 0.05$. Since the most frequently taken test was the SAT ($N = 1217$), it was used as the indicator of the math achievement as measured by conventional standardized tests for the main analysis in order to reduce the effect of missing data. Descriptive statistics for SAT scores showed that the sample consisted of students from different achievement levels with $\mu = 663.14$, and $SD = 63.27$. The distribution of students' SAT scores are displayed in Figure 1. The range of SAT scores was 350 with a maximum score of 800 and a minimum score of 450 within the sample. The range and distribution of SAT scores supported that there were students from different achievement levels in the sample.

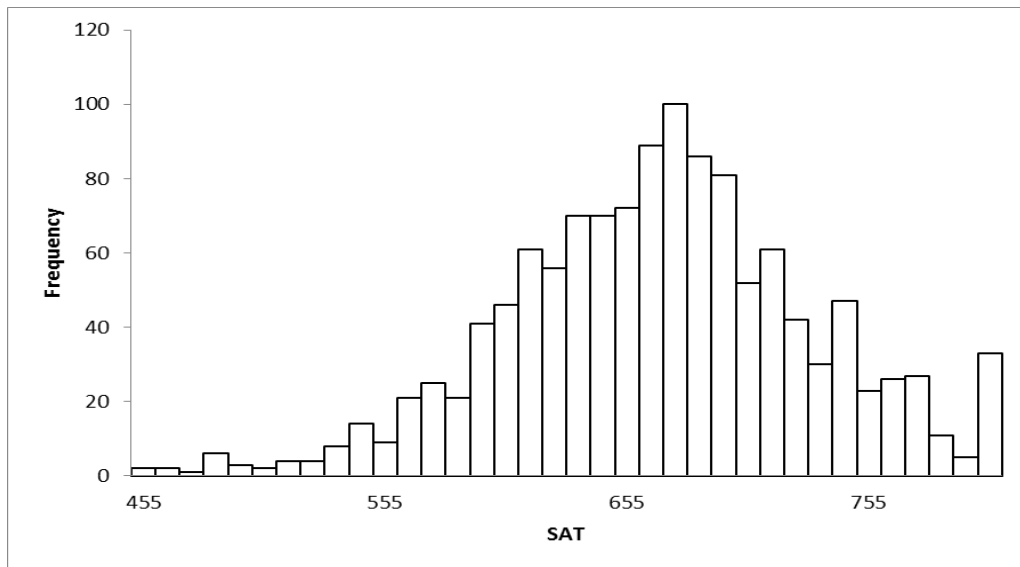


Figure 1. SAT score distribution for the sample

Findings for Sub-question II: Selecting an Indicator for Modeling Performance (i.e., MEAs)

A high positive correlation was found between students' scores on MEA-1 and 2 with a significant correlation coefficient, $r(1656) = .94, p < 0.05$. This result showed that there is no difference between the two MEA scores. The scores for MEA-1, Just-In-Time, were used in the analysis as an indicator of the mathematics achievement on the modeling task, because we had a more complete set of data for this MEA due to number of students withdrawing from the course between MEA-1 and MEA-2. MEA-1 was the third MEA in the sequence of the four MEAs that students were exposed to during their first-year experience. Descriptive statistics were provided for the MEA-1 to verify that the participant students performed at different levels on the task. The percentages of students in each level of the MEA-1 scores are presented in Table 1.

MEA-1 Score Categories	Number of Students	Percentage
0	31	1.9%
50	297	17.9%
60	722	43.6%
70	577	34.8%
80	28	1.7%
Valid	1655	100.0%
Missing	1	
Total number of students	1656	

Findings for Sub-question III: Relations between Standardized Test Performance and Modeling Performance (i.e., MEA) Scores

The major challenge in comparing SAT scores with the performance on MEA was that MEAs were done in teams. So, individual students in the same group were assigned identical MEA scores. To address a potential effect of in-group differences in SAT scores, a regression analysis with two predictors was used. In other words, whether the low/high MEA scores were attributable to low and/or high SAT scores within the teams were tested.

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, $\chi^2(1, N= 1251)= 1.509, p= .219$. More specifically, one unit increase in students' SAT scores (i.e., going from 0 to 1) resulted in a change of only 0.002 in the ordered log odds of scoring in a higher level on MEA-1 team scores. As the obtained p -value is well above the critical value of 0.05, we concluded that students' individual SAT scores do not predict their MEA-1 scores.

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 $\chi^2(2, N=174) = .561, p= .755$. One unit increase in the teams' highest SAT scores resulted in an increase of only 0.002 in the ordered log odds of scoring in a higher level on MEA-1 given the lowest SAT predictor is held constant, and one unit increase in the teams lowest SAT scores resulted in an increase of only 0.0001 in the ordered log odds of scoring in a higher level on MEA-1 given the highest SAT predictor is held constant.

Findings for Sub-question IV: Relations between Low- and High Achievers' on Traditional Tests and their Performance on Alternative Assessment (i.e., MEA)

The distribution of the students on the three percentiles as low, medium, and high achievers is presented in Table 2.

Table 2. Number of students in each achievement group on SAT

	Low	Medium	High
Range of Scores	450-620	620-700	700-800
Number of students	305	620	326

The frequencies of MEA-1 scores for the traditionally low- and high-achievers' scores are presented in Table 3. The majority (75.4%) of the SAT low-achiever group performed at the level of 60 and 70, where only a non-significant fraction of the group (1.9%) performed at the level of zero, and 20.3% performed low (at the level of 50) on the MEA-1. Combining the information from Table 1 and Table 3, one can see that only 19.3% of the students at the level of 0 on MEA-1 were SAT low-achievers, and the majority (80.7%) of the level 0 performers was comprised of SAT high- or medium-achievers. Table 1 and Table 3 also show that only 20.9% of the students at the level of 50 on MEA-1 were SAT low-achievers.

Table 3. MEA-1 category frequencies for SAT low-achiever group

	MEA-1 scores					Total
	0	50	60	70	80	
SAT low-achievers	6	62	128	102	7	305
[Percent within group]	[1.9%]	[20.3%]	[42%]	[33.4%]	[2.3%]	
SAT high-achievers	7	18	111	116	73	326
[Percent within group]	[2.1%]	[5.5%]	[34.2%]	[35.5%]	[22.3%]	

The majority (69.7 %) of the SAT high-achievers performed at the level of 60 and 70, where only a non-significant fraction of the group (2.1%) performed at the level of zero, and 5.5% performed low (at the level of 50) on the MEA-1. Table 1 and Table 3 show that 22.5% of the students at the level of 0, and 6.06% 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' mathematics achievement as measured by a conventional standardized assessment, the SAT, did not predict their performance on a modeling task, the MEA. This finding suggests that the predominant use of standardized tests to influence students toward or away from STEM professions cannot be supported, if modeling is an important indicator of capabilities needed for success in STEM professions.

Especially noteworthy was how students' scores were consistent across the three standardized tests—SAT, ACT, and ALEKS,—and across two modeling tasks—MEA-1 and MEA-2,—but not between the standardized tests and modeling tasks. This suggests that which is assessed by the conventional standardized test and that which is assessed by modeling problems are different from each other. Simply, successful performance on MEAs requires capabilities different than which is captured by conventional standardized mathematics assessments. This result provides evidence to the anecdotal and case-based claims (e.g., Lesh & Harel, 2003; Lesh & Sriraman, 2005) in the literature that there are discrepancies between students' performance on realistic modeling problems compared to their performance on traditional assessment. This result may also shed light on the problem raised by Ridgway et al. (2002) that there may be no way of assessing new educational goals such as higher-order thinking by using existing techniques—commercially available standardized tests.

Another major finding is that 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. The Iversen and Larson (2006) study documents that students who excel on traditional assessments often struggle to implement their mathematical knowledge in real-world settings, however they did not address how students who performed poorly on the traditional tests did on the modeling activities. This study addresses this issue by revealing most students with relatively low SAT scores performed as well as the students with high SAT scores on MEAs.

This study provides compelling evidence to the argument that MEAs require capabilities different than those tapped by conventional standardized test, and thus the traditional test cannot be used as predictors of students success in modeling complex problem situations, which is an essential component of STEM professions. The SAT test is considered to be keeping pace with what colleges are looking for today and measuring the skills required for success in the 21st century, however the results of this study showed that it does not capture modeling capability. The major limitation of this study was comparing individual SAT scores to team MEA scores. A major concern about this comparison was the possibility of situations where one student with a high SAT score was leading a team, and other students with low SAT scores were just the followers during the modeling process, or vice versa. In order to overcome this limitation, the lowest and highest SAT scores from each team were added to the regression model as predictors of team MEA scores. This kind of analysis ameliorated the limitation. The population sample of the study was a sample of convenience. Therefore, the minor limitation is that the sample was not necessarily a random or representative sample of the total population of college students in U.S. Nonetheless, the sample was the universe of this particular university, and the diversity of student scores on the standardized tests showed the variety of students at the university.

References

- Accreditation Board for Engineering Technology, Inc. (ABET). (2010). *Criteria for accrediting engineering programs: Effective for evaluations during the 2010–2011 accreditation cycle*. Baltimore, MD.
- Baron, J., & Norman, M. F. (1992). SATs achievement tests, and high-school class rank as predictors of college performance. *Educational and Psychological Measurement*, 52(4), 1047-1047.
- Carpenter, J., & Hanna, R. E. (2006, June). *Predicting student preparedness in calculus*. Paper presented at the American Society for Engineering Education Annual Conference, Chicago, IL.
- 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.
- Cizek, G. J. (2010). *Translating standards into assessments: The opportunities and challenges of a common core*. Chapel Hill, NC : University of North Carolina at Chapel Hill.
- 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., Follman, D., Imbrie, P. K., Zawojewski, J., Capobianco, B., & Hjalmanson, M. (2004, June). *Model eliciting activities: An in-class approach to improving interest and persistence of women in engineering*. Paper presented at the American Society for Engineering Educating Conference, Salt Lake City, UT.

- 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.
- Diefes-Dux, H. A., Hjalmarson, M., Zawojewski, J. S., & Bowman, K. (2006). Quantifying aluminum crystal size part 1: The model-eliciting activity. *Journal of STEM Education: Innovations and Research*, 7(1), 51-63.
- Diefes-Dux, H. A., Zawojewski, J. S., & Hjalmarson, M. A. (2010). Using educational research in the design of evaluation tools for open-ended problems. *International Journal of Engineering Education Special Edition* (26)4, 807-819.
- Diefes-Dux, H. A., Zawojewski, J.S., Hjalmarson, M., & Cardella, M. (2012). A framework for analyzing feedback in a formative assessment system for mathematical modeling problems. *Journal of Engineering Education*, 101(2), 375-406.
- Educational Testing Service (1948-1990). *College Board Scholastic Aptitude Test*. Princeton, NJ: Author.
- Falmagne, J. C., & Doignon, J. P. (2011). *Learning spaces*. Berlin: Springer.
- 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.
- Iversen, S. M., & Larson, C. J. (2006). Simple thinking using complex math vs. complex thinking using simple math: A study using model eliciting activities to compare students' abilities in standardized tests to their modeling abilities. *ZDM*, 38(3), 281-292.
- 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. A. (2003). How mathematizing reality is different from realizing mathematics. *Mathematical Modeling: A Way of Life ICTMA11*, 37-52.
- 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. (2000). Symbolizing, communicating, and mathematizing: Key components of models and modeling. In P. Cobb & E. Yackel (Eds.), *Symbolizing and Communicating in Mathematics Classroom: Perspectives on discourse, tools, and instructional design* (pp. 364-381). 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., & Kelly, A. E. (1997). Teachers' evolving conceptions of one-to-one tutoring: A three-tiered teaching experiment. *Journal for Research in Mathematics Education*, 28(4), 398-430.
- 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.
- Marklein, M. B. (March 19, 2007). All four-year U.S. colleges now accept ACT test. *USA Today*. Retrieved from <http://www.usatoday.com/>.
- Miller, S. (1995). *An American imperative*. New Haven, CT: Yale University Press.

- Moore, T., & Diefes-Dux, H. (2004, October). *Developing model-eliciting activities for undergraduate students based on advanced engineering content*. Paper presented at the 34th ASEE/IEEE Frontiers in Education Conference. Savannah, GA.
- Myers, R. S., & Pyles, M. R. (1992, November). *Relationships among high school grades, ACT test scores and college grades*, Paper presented at the Mid-South Educational Research Association, Knoxville, TN.
- National Academy of Engineers (2005). *Educating the engineer of 2020: Adapting engineering education to the new century*. Washington, D.C.: National Academy of Engineers.
- 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 (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics (2013). *Formative Assessment*. Reston, VA: Author.
- National Council of Teachers of Mathematics (2013). *Supporting the Common Core State Standards for Mathematics*. Retrieved from <http://www.nctm.org/ccssmposition/>
- National Science Board (2000). *Science and engineering indicators 2000* (NSB-00-1). U.S. Government Printing Office, Washington, DC.
- National Science Foundation, National Center for Science and Engineering Statistics (2013). *Women, Minorities, and Persons with Disabilities in Science and Engineering*. Arlington, VA.
- Niss, M. (1999). Aspects of the nature and state of research in mathematics education. *Educational Studies in Mathematics*, 40(1), 1-24.
- Reddy, A., & Harper, M. (2013). Mathematics placement at the University of Illinois. *PRIMUS*, 23(8), 683-702.
- Ridgway, J., Zawojewski, J. S., & Hoover, M. N. (2000). Problematising evidence-based policy and practice. *Evaluation & Research in Education*, 14(3-4), 181-192.
- Schoenfeld, A. H. (2002). Making mathematics work for all children: Issues of standards, testing, and equity. *Educational researcher*, 31(1), 13-25.
- Senk, S. L., Thompson, D. R. (2003). *Standards-based school mathematics curricula: What are they? What do students learn?* Mahwah, NJ: Lawrence Erlbaum Associates.
- 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. A., & Diefes-Dux, H. A. (2013). *A teaching assistant training protocol for improving feedback on open-ended engineering problems in large classes*. Paper presented at the 120th ASEE Annual Conference and Exposition, Atlanta, GA.
- Torra, V., Domingo-Ferrer, J., Mateo-Sanz, J. M., & Ng, M. (2006). Regression for ordinal variables without underlying continuous variables. *Information Sciences*, 176(4), 465-474.
- 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.