Dual Measures of Mathematical Modeling for Engineering and Other STEM Undergraduates

Jennifer A. Czocher, Kathleen Melhuish, Sindura Subanemy Kandasamy & Elizabeth Roan

ONLINE FIRST

International Journal of Research in Undergraduate Mathematics Education

ISSN 2198-9745

Int. J. Res. Undergrad. Math. Ed.
DOI 10.1007/s40753-020-00124-7

Springer

International Journal of Research in Undergraduate Mathematics Education
https://doi.org/10.1007/s40753-020-00124-7

Dual Measures of Mathematical Modeling for Engineering and Other STEM Undergraduates

Jennifer A. Czocher[1] · **Kathleen Melhuish**[1] · **Sindura Subanemy Kandasamy**[1] · **Elizabeth Roan**[1]

Accepted: 23 October 2020/Published online: 04 January 2021

Abstract
This paper addresses two aspects of integrating mathematics education with engineering education that may address persistence of engineering majors (and STEM majors more broadly): an emphasis on modeling as a vehicle for more authentic learning activity (Niss et al. 2007), and the need for measures that can support academic units' efforts to collect local data about student attainment of program goals. In this paper, we contribute: (1) a measure for modeling self-efficacy and its corresponding design process; (2) a measure for modeling competency and its corresponding design process; (3) a preliminary analysis of the relationship between modeling competency and self-efficacy. We argue that such instruments address a genuine need of engineering departments (as well as STEM education researchers) to have a means for collecting local data on students' modeling self-efficacy and competency.

Keywords Mathematical modeling · Undergraduates · STEM majors · Assessment · Modeling competencies · Self-efficacy

✉ Jennifer A. Czocher
czocher.1@txstate.edu

Kathleen Melhuish
kmm335@txstate.edu

Sindura Subanemy Kandasamy
s_k252@txstate.edu

Elizabeth Roan
eaw109@txstate.edu

[1] Texas State University, (Mathematics), San Marcos, Texas, USA

Introduction

According to the Washington Accords held by the International Engineering Alliance (2014), mathematical modeling of problems in engineering is one of the essential skills desired of graduates in engineering. This is consistent with other governing bodies, such as the European Network for Accreditation of Engineering Education (2018). That is to say, engineering graduates must be able to fluently apply mathematical knowledge to complex engineering problems. Mathematical modeling serves an essential role in providing engineering students' authentic experiences which leverage the engineering design theories, scientific inquiry, technological literacy, and mathematical thinking as they are used by the greater engineering community (Kelley and Knowles 2016). Modeling skills are paramount for an increasingly technologically-literate job market as well as for the societal problems whose solutions have global consequences – problems that today's students, are interested in solving (Su et al. 2009; Eccles and Wang 2016). Furthermore, engineering students also value instruction that integrates the mathematics they learn with real-world applications (Güner 2013). More, empirical studies suggest positive gains for students who are exposed to mathematical modeling in terms of their self-efficacy and robustness of mathematical knowledge (Czocher 2017; Czocher et al. 2019; Lesh et al. 2000; Rasmussen and Kwon 2007; Sokolowski 2015). We recently synthesized the literature to argue that mathematical modeling, as a pedagogical approach, has the potential to increase student interest, proficiency in mathematics, and self-efficacy(Czocher et al. 2019). Moreover, these factors are severally and jointly positively associated with persistence in STEM fields, including engineering. For these reasons, many faculty have begun revising their course materials and many academic units in STEM have made tangible changes to their programs to incorporate mathematical modeling as a source of motivation and as a learning outcome in its own right (Chiel et al. 2017; Moore et al. 2013; Yildirim et al. 2010).Since mathematical modeling activities can be small-scale and can potentially be drawn from narrow contexts, they can be easily integrated into existing engineering and mathematics curricula (Kertil and Gurel 2016; Hallström and Schönborn 2019). This is especially the case for advanced mathematics courses like differential equations (Czocher 2017; Liu and Raghavan 2009; Pennel 2009; Rasmussen and Kwon 2007) and linear algebra (Dominguez-Garcia et al. 2016; Trigueros and Possani 2013), topics that naturally lend themselves to mathematical modeling and are often challenging courses for engineering majors to maintain interest and to learn from but nevertheless act as gatekeepers for many engineering practices.

Taken together, this body of research points to the importance of modeling experiences for engineering students learning mathematics, and the potential for instructional innovations that incorporate modeling to positively impact their academic careers. From the literature, we know that minor but meaningful improvements to engineering students' mathematical education may have outsized effects on retention due to improvements in both understanding and affective constructs (self-efficacy; self-confidence). As engineering programs and mathematics classrooms increasingly incorporate mathematical modeling as a learning outcome, there is a need for faculty, academic units, and programs to have ways to assess their students' mathematical modeling competence and their self-efficacy in modeling. Evaluations that focus on final grades or student perceptions provide little insight into the direct impact on engineering students' growth in relation to modeling. Validated, reliable, and carefully linked assessment is vital for measuring learning outcomes and demonstrating the efficacy of innovative pedagogies. It is also vital for continual refinement of classroom and programmatic improvements and for synthesizing empirical findings about these outcomes. Collecting informative data can be a high-stakes endeavor for academic units to demonstrate meeting credentialing boards' criteria for engineering programs. At times, a stakeholder such as a departmental task force, may not know what should be evaluated or how (Hoey and Nault 2008). Such barriers are superable only when faculty have trust in the use of assessment – trust, that can be provided through "sound methodological frameworks and robust instrumentation" (p. 187).

In a survey of mathematics education research literature from leading outlets for work in mathematical modeling, Frejd (2013) found that the majority of modeling assessments (at any level) were not grounded in theory but rather were based on ad hoc constructions, personal experience, or small-scale studies of student work. More generally, even when assessments exist, stakeholders in education may be reluctant to adopt them or trust them enough to base policy decisions on them. Further, many of these existing assessments are not suitable for use with advanced mathematics or within engineering education. Prior to our work, there as a lack of validated, reliable instruments for assessing growth of students' modeling skills situated in the advanced mathematics taken by undergraduate engineering majors. Despite the clear focus of mathematics and engineering educators, these fields face a key impediment in exploring efficacy of their instructional innovations regarding mathematical modeling.

Taken together, we view these concerns as a call for a validated, reliable instrument capable of measuring gains associated with instructional interventions centered on mathematical modeling. Such instruments are particularly vital for engineering programs. In this paper, we share design principles and empirical evidence to conduct a validation study for a pair of modeling assessments situated in differential equations:

1. The Modeling Competency Questionnaire (MCQ) which assesses modeling competencies.
2. The Mathematical Modeling Self-Efficacy Measure (MSE) targeting self-efficacy for carrying out modeling competencies.

We document their development and their properties. We will offer evidence of their validity and reliability, based in multiple rounds of empirical testing and describing how the instruments' development were grounded in the extant literature on teaching and learning of mathematical modeling for STEM majors. As another source of validity, and to explore and confirm the utility of the instruments, we also address the following research questions:

1. Do the instruments reflect an association between self-efficacy and competency in mathematical modeling?
2. Do the instruments detect significant differences in measured outcomes based on academic major, gender, proxies for mathematical knowledge, or previous experience participating in modeling competitions?

Thus, our contribution is to employ theory and empirical findings to validate the much-needed instruments and thereby increase confidence in their utility.

Theoretical Framing: Operationalized Constructs and Logic Model

In this section, we provide theoretical framing for and backing from educational literature to situate the self-efficacy and competency instruments as well as their use. We first present a literature-based argument for the importance of modeling, then a framework for modeling competencies and an operationalization of self-efficacy that underscores our instrument development process. We conclude with considerations about the intended use of instruments of this nature in engineering education.

The Important Role of Modeling for Engineering and Other STEM Majors

Previously, we argued for the value of mathematical modeling as a pedagogical approach to teaching mathematics in terms of its potential to ameliorate STEM including engineering attrition rates (Czocher et al. 2019). Based on our review of empirical results from extant research literature, we present the following inferences as working assumptions:

- Modeling helps students see why and how mathematics is relevant to their coursework, their careers, and their daily lives through interesting problems to solve (e.g., Bliss et al. 2016).
- Students are more interested in learning content relevant to them (e.g., Kim et al. 2015), and so modeling indirectly increases their motivation to learn.
- Modeling can help students learn mathematics content, thereby increasing mathematical proficiency (Sokolowski 2015; Young et al. 2011).
- Modeling enables student-centered pedagogies, which in turn lead to gains in self-efficacy(e.g., Schukajlow et al. 2012).
- Taken together, these strengths of modeling are precursors to persistence in mathematics, which is requisite for persistence across STEM majors.

As we argued previously, there is substantial accumulated evidence that mathematical modeling is positively associated with mathematics proficiency, self-efficacy, and answers for students the question “why is <mathematics topic> relevant to me?” Measures of these three domains are positively associated with persistence in mathematics, ultimately leading to persistence of STEM majors. Self-efficacy is an important construct for engineering students in particular since it correlates to mathematics achievement (Loo and Choy 2013) and their decisions to persist (Jones et al. 2010). We view innovations in mathematics instruction that center on mathematical modeling as particularly advantageous to engineering undergraduates since the attrition rates for engineering majors were historically around 50% from the 1950s through the 2010s (Geisinger and Rajraman 2013). It is thus important to be able to measure growth in modeling skills and self-efficacy to do mathematical modeling that may be fostered by such innovations.

Operationalization of Modeling and Self-Efficacy

For this project, we adopt a view of mathematical modeling as a process of rendering a non-mathematical problem about a real-world phenomenon of interest as a well-posed mathematical problem to be solved. We focus on the cognitive activities that facilitate the process (Kaiser 2017) and utilize a mathematical modeling cycle (MMC), such as the one in Fig. 1, to operationalize these activities as skills or competencies (Blum and Leiss 2007; Czocher 2016). There are many modeling cycles available in the educational literature, including some well-known ones from engineering education (e.g., Dym 2004). We chose to work with the version popularized by Blum and Leiss (2007) because it has been empirically tested and expanded (Czocher 2016) as we outline below.

According to the MMC, mathematical modeling is an iterative and cyclical process. It begins with.

specifying a problem, that is, *understanding* what the real-world problem should be or even identifying what aspect of a real-world phenomenon is worth inquiry. Then, the modeler must separate the relevant factors from ones that can be safely ignored. These *structuring* choices often involve judgment in making assumptions reflecting the modeler's mathematical and non-mathematical knowledge as well as their (or their stakeholder's) values. At times, progress through the process may be impeded by these choices. The modeler may introduce assumptions that warrant mathematical techniques they do not command or identify variables that make the resulting model overly complex. Sometimes, the modeler may greatly simplify the problem by ignoring a variable to the detriment of overall accuracy. At other times, deciding not to consider variables may not have too great a cost to overall accuracy of predictions and usefully simplify the problem at hand (e.g., ignoring wind resistance for an object in a short free fall). The conditions and variables need to be *mathematized*, that is, presented in terms of conventional mathematical representations. The representations may be equations, graphs, tables, or algorithms. The representations signify relationships among quantities – how they vary and covary and the mathematical rules they obey. This part of the process poses a mathematical question to be answered. The resulting model must be analyzed to provide a mathematical answer. Typically, *working mathematically* is

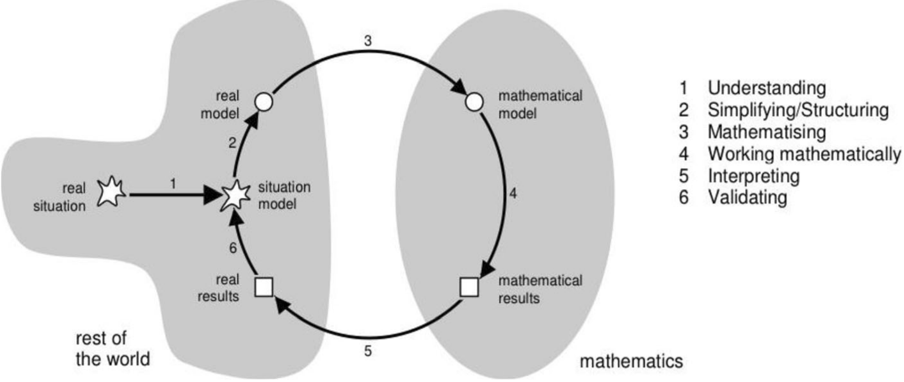

Fig. 1 Mathematical modeling cycle, as elaborated by Blum and Leiss (2007)

handled in mathematics classes because solving mathematical problems is their purview. For example, learning techniques to solve a first-order, homogeneous, linear differential equation would be taught in a typical course on ODEs. The mathematical results must be *interpreted* in context so that the results and the model itself might be *validated*. Validating involves checking that the model is representative of the situation and articulating its limitations.

Collectively, these activities are typically referred to as modeling competencies (Maaß 2006). They reflect the cognitive and metacognitive complexity of blending mathematical knowledge with real-world knowledge (Czocher 2016; Fauconnier 2001; Stillman and Galbraith 1998; Stillman 2000). Developing these competencies is challenging for students. They often struggle to define mathematical problems from real-world situations and to validate the models they generate because there can be an overwhelming number of considerations. For these reasons, it is also important to consider students' self-efficacy related to modeling as their skills grow.

We then follow Hackett and Betz (1989); Betz and Hackett (1983); Bandura (2006) in operationalizing self-efficacy*about* a task as an individual's perceived capacity to successfully carry out that task. In Czocher et al. (2019), we introduced the specialized construct *mathematical modeling self-efficacy*(MSE) to mean "an individual's perceived capability to carry out the interrelated activities that make up mathematical modelling" (p. 13).

The Role of Modeling Measures in Engineering Education

Because this paper focuses on instrumentation, our last theoretical consideration is in relation to intended use. We situate our instruments as meeting a research and engineering program need for continual improvement. Continuous improvement of programs is emphasized by organizations across nations such as in the United States' ABET (see ABET Engineering accreditation Comission 2018-2019) or Sweden's CDIO standards (see Malmqvist et al. 2006). Continuous improvement is adapted from quality management and emphasizes improvement integrated with assessment through cycles of *planning* via identifying a need for change, *doing* or implementing the change, *checking* to see if the original goals are achieved, and *acting* to continue to implement the innovation or make changes as a result of data collected (Temponi 2005). In engineering education (and higher education in general), it is essential to collect data on student outcomes as part of continuous improvement (McGourty et al. 2013). In this paper, we introduce instruments that can serve to meet particular *goals* related to engineering education: incorporating of modeling, and a means to collect student outcome data through related assessments. Such data collection serves a key role in the continuous improvement of engineering education for students and can directly impact their education through identifying the impact of didactical changes focused on modeling.

Modeling Self Efficacy Questionnaire

In this section, we address the first goal of this paper: presenting the modeling self-efficacy instrument. We share the development process along with evidence of the validity and reliability of the instrument within the context of its development cycles. The purpose of the instrument is to measure growth in learners' mathematical modeling self-efficacy. It is intended to be used as a pre/post pair to provide feedback on educational interventions arising from educational research or pedagogical innovations whose desired outcomes are modeling competencies. Throughout this project, we chose to focus on the domain of differential equations because it is a standard subject required of engineering programs and is a natural location to incorporate mathematical modeling alongside mathematical content that was developed in response to historical, real-world modeling problems. For these reasons, our target population is students who are enrolled in or have previously taken differential equations (or some similar course focusing on relationships among quantities and their derivatives).

Instrument Development

We designed a situation-specific survey instrument, with explicit reference to competencies of mathematical modeling and theory of measurement of self-efficacy(Bandura 2006). The items asked students to report self-confidence on modeling competencies known to be instrumental to successfully carrying out mathematical modeling: making assumptions, estimating parameters, identifying variables, mathematizing, validating, establishing limitations, working mathematically, and communicating. The initial instrument had six self-efficacy statements and was subsequently modified to include all eight competency statements (Table 1). Each item was a Likert-type statement targeting the modeling competencies on a 100-point scale, with 10-unit intervals, from 0 ('Cannot do'), through intermediate degrees of assurance 50 ('Moderately certain can do'), to complete assurance 100 ('Highly certain can do'). The Modeling Self Efficacy (MSE) instrument has been tested on four occasions, including the pilot run. On each occasion, samples were drawn from an annual, international mathematical modeling competition whose focus is on modeling with differential equations.

Round 1 The initial questionnaire consisted of six self-efficacy statements. For the 38 competitors in the questionnaire prior to the October 2017 competition, we conducted a principal component analysis (Abdi and Williams 2010). The analysis extracted one underlying dimension accounting for 62.49% of the variance in scores. Cronbach's alpha (Cronbach 1951) was then calculated on the set of six items on the pre-test, $\alpha = 0.822$, reflecting a high degree of internal consistency. For the 21 participants who completed both the pre- and post-competition questionnaires, a matched pairs *t*-test revealed gains in self-efficacy(Czocher and Kandasamy 2018).

Round 2 In order to address a focal goal of the competition, we added a self-efficacy statement regarding *establishing limitations*(a *validating* competency, see Table 1, Item 6). We also made minor adjustments to improve clarity of the statements. A principal component analysis was conducted for the set of pre-competition responses ($n = 274$) to explore dimensionality. The analysis extracted one underlying dimension accounting for 67.12% of the variance in scores. All items were pair-wise correlated with correlation coefficients greater than 0.4 and $p < 0.001$, respectively. Cronbach's α for the seven items was $\alpha = 0.917$, reflecting a high degree of internal consistency**.** Overall, student gain in modeling self-efficacy was statistically significant ($t(92) = -6.663$, $p < 0.001$). This difference had a moderate effect-size, $d = 0.545$. A more nuanced

Table 1 Final MSE instrument

Rate your level of confidence by recording a number from 0 to 100 using the scale given below: 0 10 20 30 40 50 60 70 80 90 100 — Cannot do at all (0), Moderately can do (40–60), Highly certain can do (70–100)	Competency
1. Create a differential equation model for the spread of smart home appliances in the United States during the twenty-first century.	Mathematize
2. In (1) identify the important variables leading to a reasonably accurate prediction.	Identify variables
3. In (1) make simplifying assumptions to reduce the number of important variables.	Make assumptions
4. In (1) select an appropriate numerical, graphical, or analytic technique to solve the resulting differential equation	Working Mathematically
5. In (1) consult appropriate resources to check whether your model was reasonable.	Validate
6. In (1) list the real-life and mathematical limitations of your model.	List limitations
7. In (1) create a short presentation to convince a smart appliance manufacturer that they could rely on your model to develop their business plan.	Communicate findings
8. Given a differential equation which describes the rate of formation of material A, $A'(t) = \alpha A(t)^{\beta}$ and a data set of observations for time, t, amount of material A at each time t, you could estimate the parameters α and β.	Estimate parameters

analysis revealed that women's gains were greater than men's and that gains for students who had not previously taken differential equations were greater than for those who had (see Czocher et al. 2019, for more details). The fact that an intervention designed to build modeling self-efficacy showed gains contributes to validity of the MSE scale.

Round 3 We added a self-efficacy statement regarding *working mathematically*(see Table 1, Item 4) which initially was left out because the focus was on the complementary competencies of modeling. A principal component analysis was conducted for the full set of pre-competition data ($n = 198$) to explore dimensionality. This analysis extracted one underlying dimension accounting for 61.463% of the variance in scores. Pair-wise item correlations additionally revealed that all items were pair-wise correlated with correlation coefficients greater than 0.3 and $p < 0.001$, respectively. Cronbach's alpha was then calculated on the set of eight items on the pre-test, $\alpha = 0.908$, reflecting a high degree of internal consistency. Collectively the students who took both the pre- and post-competition questionnaire showed gains in Modeling Self-Efficacy from before to after the competition ($t = 4.202, df = 51, p < 0.001$).

Summary of Field Testing

These analyses along with its creation rooted in theory of mathematical modeling competencies suggest the Modeling Self Efficacy (MSE) scale is an internally consistent, unidimensional instrument with high face and construct validity. Therefore, the responses to the items of this instrument can be summed to measure individuals' self-efficacy in modeling, overall. Over time, the properties of the MSE scale have remained stable and it is capable of measuring gains, suggesting both reliability for measuring the modeling self-efficacy construct and utility towards that purpose.

The Modeling Competency Questionnaire

One of the most widely conducted studies and therefore accepted forms of evidence for documenting benefit of an educational intervention is a pre−/post- with comparison group study. Thus, our purpose in developing the Modeling Competency Questionnaire (MCQ) was to facilitate generation of the kinds of evidence valued by researchers, instructors and program administrators, namely, to show gains not only from before to after a modeling-based instructional intervention but also in comparison to a control group. For these reasons, we sought to develop a pair of parallel assessments that could be used to conduct quasi-experimental studies of educational innovations for undergraduate STEM majors involving mathematical modeling. That is, a student could take one version as a pre-test and the parallel version as a post-test.

Frejd (2013) estimated that about one-third of modeling assessments were written multiple-choice tests based on Haines et al. (2000). Items in that assessment were designed to target a single competence within the modeling process (e.g., asking clarifying questions, identifying variables). The distractor options were created to be either irrelevant to the construction of a model or to consider only the real-world constraints without mathematical constraints. The "best" answer choice considered both real-world constraints and relevant mathematics. Despite the promise of the instrument, there are a few substantive critiques. First, the question set was tested using Rasch analysis on a sample of 39 students and two sets of six items, so its broader properties are unknown. Because of this, there is some reason to question whether the parallel forms were indeed comparable in difficulty and in content. Second, two intervening decades of research advances have provided a wealth of information about plausible distractor choices that align with students' tendencies. Thus, it is possible to improve validity of the instrument by drawing on that research to generate distractors. Third, there have been further advances in psychometric methods allowing robust treatment of dichotomously keyed items. Our instrument addresses these critiques.

Challenges to Assessing Modeling

One of the most challenging aspects of assessment design in this domain is managing the many trade-offs: between specificity and generality, between authenticity and manageability, between openness to student thinking and facility in grading. Since success in mathematical modeling is dependent upon both mathematical knowledge and nonmathematical knowledge about the scenario, any assessment is necessarily bound by the specific contexts the items are drawn from and necessarily brings with it the uncertainty of students' prior knowledge. Thus, there is a natural desire in assessment to reduce noise in the measurement that could be introduced by many contexts; however, an assessment drawn from a single context would not be valid for measuring modeling competence because it would measure facility with only that specific situation. The same could be said when considering the mathematical dimension – a student may have a deep understanding of the scientific principles at stake in a context but lack mathematical knowledge or capacity to apply it to successfully model the scenario. Decisions at each stage of the modeling process depend on the decisions that were made earlier in the process. Because competencies are interconnected, assessment design should target the reasoning underscoring student decision-making as well as the modeling decisions. That is, an assessment of the extent to which an individual has developed a competency should acknowledge that a student's choices may diverge from a normatively correct answer but also recognize their justifications for their choices (Czocher 2019). Said plainly, an assessment in modeling treats, at minimum, three interrelated dimensions which are context- and student-dependent: real-world knowledge, mathematical knowledge, and capacity to relate the two. A valid assessment, then, would sample students' modeling competence across a range of scenarios and a range of mathematical content. Accounting for students' reasoning is challenging to do with a multiple-choice instrument, but it is possible when distractors are rooted in students' ways of reasoning about modeling scenarios. Moreover, because the assessment's utility is for evaluating instructional interventions or innovations commenced by individual faculty, programs, academic units, or researchers, the intent is not to furnish the assessment of knowledge or skills directly to students. This kind of *local data* is indispensable for understanding how or why instruction succeeds (or fails) (see, for example, Bressoud and Rasmussen 2015) and can be based on simply-generated and consistent measures of growth in students' skills that are rooted in theory and empirical studies of students' mathematical modeling.

Our approach to multiple-choice item construction adhered to a set of constraints. First, we insisted that the scenarios and problems were relevant and authentic. We operationalized this constraint by drawing from source material students would encounter in their studies (e.g., radioactive decay) or in public, contemporary discourse (e.g., recycling). Second, we sought to phrase question stems to appeal to multiple domains of knowledge, including mathematics, science, engineering, and "everyday commonsense." Third, we drafted question stems that would target aspects of competencies via alignment with specific indicators of the modeling competencies (see Czocher 2016). Finally, we constructed distractors based on previous research studies treating the kinds of concerns and justifications that students exhibit while modeling. An overview of the development process is depicted in Fig. 2.

Item and Scale Development

We developed a pool of 118 multiple choice questions (MCQs) targeting the MMC competencies drawn from 9 real-worldscenarios and some selections from the original Haines et al. (2000) items. We did not include *working mathematically*, since analysis is typically the focus of mathematics coursework. The real-world scenarios were drawn from research and educational materials (e.g., GAIMME report; textbooks; published

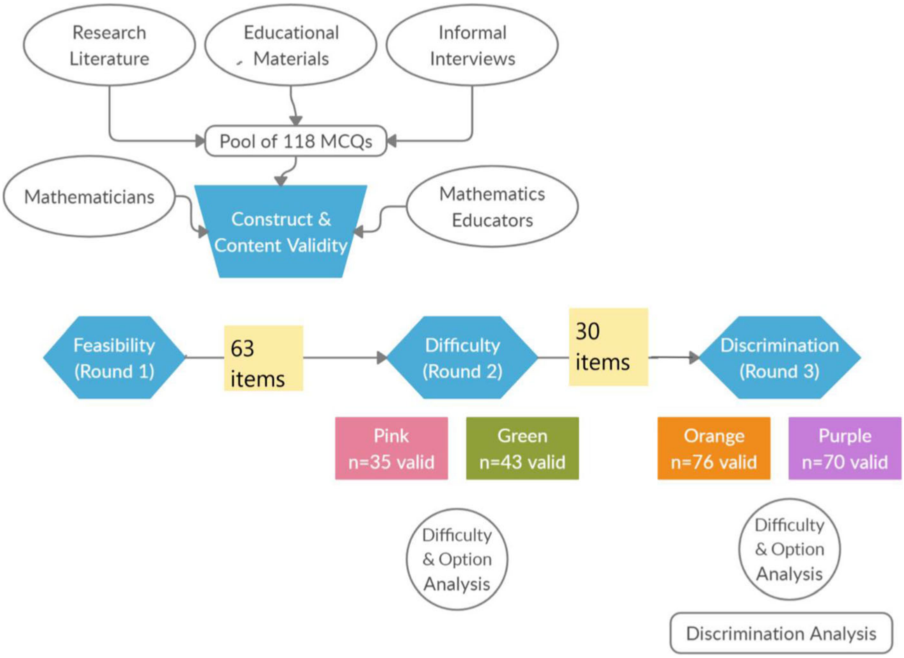

Fig. 2 Overview of MCQ instrument development, made in Creately

research; faculty syllabi) that are appropriate to STEM undergraduates who have completed integral calculus. We sought scenarios that treated prevalent social issues, involved situations in the sciences where differential equations might be used, or were suggested during interviews with STEM professors. The contexts were (briefly): a fracking wastewater lagoon, a municipal recycling program, radioactive decay, kinematics, moth flight, population growth, carrying capacity for a population of white-tailed deer, disease transmission, and several items from various contexts developed by Haines and colleagues. A list of item titles and their competence targets is in Supplementary 1. Mathematical content ranged across arithmetic, algebra, calculus, and systems of ordinary differential equations. We drafted MCQs from each scenario, striking a balance between information provided in the scenario set-up (so that the problems the MCQs addressed were situated) and readability (so that multiple MCQ stems could follow, cognitively, quickly from the set-up). A variety of question stems were used (e.g., select the most/best/least; indicate the choice consistent with the assumptions) and responses were developed to have a single "best" answer with four distractors at varying degrees of reasonability. For example, reasonability for a *structuring* MCQ might query (un)helpful assumptions to make. Sample MCQs are in Table 2 (scenario set-ups omitted due to space constraints).

Content and Construct Validity To establish content validity, we invited two mathematicians who regularly teach differential equations to mathematics and engineering majors to examine the questions for readability, adequacy of answer choices, correctness of mathematics, and appropriateness for their students. To establish construct validity, we invited three mathematics educators with expertise in mathematical modeling to evaluate the items for readability of the questions, adequacy of the answer

Table 2 Sample MCQs targeting aspects of understanding (from Haines et al. 2000.) and mathematizing, respectively

Understanding (clarifying questions)	Mathematizing (Choosing a representation)
Consider the real-world problem (do not try to solve it!): *What is the best size for stroller wheels?* Which one of the following clarifying questions most addresses the smoothness of the ride as felt by the child?	Given all of the assumptions below, which equation **best** models the growth of a human population? 1. The human birth rate is proportional to the population present, 2. There are sufficient resources (e.g., space and ample food) for the population to thrive, given 3. People die of old age and also prematurely, for example, due to malnutrition or inadequate medical supplies. Deaths also occur due to unnatural causes such as communicable disease and violent crimes. 4. Deaths are proportional to the number of two-party interactions 5. k_1 and k_2 are the proportionality constants for birth rate and death rate respectively
a) Does the stroller have three or four wheels?	a. $\frac{dP}{dt} = k_1P - k_2\frac{P^2}{2}$
b) What is the distance between the front and the back wheels? *	b. $\frac{\mathrm{dP}}{\mathrm{dt}} = k_1P + k_2\frac{P(P-1)}{2}$
c) Is the seat padded?	c. $\frac{dP}{dt} = k_1P - k_2\frac{P}{2}$
d) How old is the child?	d. $\frac{dP}{dt} = k_1P - k_2\frac{P(P-1)}{2}$ *
e) What is the surface material that the stroller will ride over?	e. $\frac{dP}{dt} = k_1P - k_2\frac{(P-1)}{}2$

choices, and whether the items were appropriately categorized in terms of the competency it was intended to target. We implemented revisions and suggestions, eliminating MCQs that failed to be correct or sensible or that duplicated other items. The remaining 59 items were tested for feasibility.

Round 1 (Feasibility) The 59 items were sorted onto 3 forms to balance competencies and tested for feasibility with 14 STEM undergraduates enrolled in courses with differential equations as a pre- or -co-requisite. Each item was followed by a set of feedback questions about readability of the question and answer choices, why they chose the answer they did, and whether they would have chosen a different answer had it been available. Each scenario was followed by a set of feedback questions about authenticity and believability of the real-world scenario, information about the scenario that would have helped them to answer the question, and mathematics knowledge that would have helped them answer the question. We examined whether students' responses were conceivably correct (justifiable) based on the reasoning provided. We revised the MCQs so that justifiable reasons would either no longer apply or else used the student reasoning to enrich distractors. We also amended scenario set-ups to include additional relevant (but not necessarily relevant-to-model-construction) information where multiple individuals indicated that there was missing information. Where possible, we left ambiguity in the scenario intact, but removed ambiguity from the answer selections.

Round 2 (Difficulty) Sixty-three items, including 59 from Round 1 (8 of which were drawn from Haines & Crouch, et al.'s previous assessments), were sorted onto two forms (Pink and Green) to balance competencies and scenarios. A total of 78 undergraduate STEM majors enrolled in courses requiring differential equations or modeling at a large university in the United States tested the items, of whom 35 completed Pink and 43 completed Green. The mean item difficulty revealed that most items (76%) were moderately difficult ($0.20 < p < 0.70$). Sixteen items performed outside this range and were flagged for restructuring. A difficulty advantage analysis revealed that nine items were too advantageous for students who took differential equations. None of these items were used in Round 3 (below). Six items were advantageous for those who had not taken differential equations. Four of these were used in Round 3 because they satisfied other constraints (see below). To analyze distractor efficiency, we calculated the proportion of students who selected each distractor. Of 253 distractors (62 items had 4 distractors and 1 item had 5), a majority were selected by at least 5% of respondents. In 17 of the items, distractors were selected more often than the keyed option. These were flagged as items with potential to be discriminating items among students with varying abilities or as potentially requiring restructuring.

After restructuring items flagged by the option analysis, 30 items were selected for Round 3 on the following basis: (i) item difficulty was between $0.20 < p < 0.70$ (ii) items should be sorted onto two forms (Orange and Purple) with comparable total difficulty (iii) each form should contain the same number of items for each competency (iv) advantage to students having completed differential equations should not be too large. There were 28 items satisfying all criteria, so we selected an additional two items satisfying (ii), (iii), and (iv) but with difficulty $p = 0.19$. Each form balanced 4 items targeting *Mathematizing*, 4 items targeting *Validating*, 4 items targeting *Structuring*, 2 items targeting *Understanding*, and 1 item targeting *Interpreting*.

Round 3 (Discrimination) The Orange and Purple forms were administered to the same sample from the international competition used in Round 4 field testing of the MSE. Of 226 students responding to the pre-event survey, 70 provided a valid response to the Purple form and 76 provided a valid response to the Orange form. We defined valid to mean fewer than 5 skipped MCQs from the end of the assessment. That is, if a respondent answered 10 of 15 questions, we counted the response as invalid if the 5 skipped questions all appeared at the end of the assessment because we interpreted this as not completing the assessment. In contrast, if the 5 skipped questions were dispersed throughout, we interpreted the student as not knowing the answer and skipping the item. Due to a survey platform error, Recycling 4 (structuring competency) and Population 4 (validating competency) were omitted from the difficulty and discrimination analyses.

The overall mean score for the Purple form was 5.4 ($SD = 2.46$) and the overall mean score for the Orange form was 5.72 ($SD = 2.31$). An independent samples *t*-test ($t = -0.811$, $df = 144$, $p = 0.419$) confirmed that there was no significant difference in mean score across the forms implying the forms were equally difficult for this sample as well as that the two groups' responses had equal variance (Hartley's F test: $F = 1.13$,

$p = 0.30$). The mean item difficulty for the Purple from was 0.39 ($SD = 0.176$) ranging from $p = 0.17$ to $p = 0.69$. The mean item difficulty for the Orange form was 0.41 ($SD = 0.165$) ranging from $p = 0.14$ to $p = 0.66$. Across both forms, four items performed at or worse than chance. We further examined the difference in proportion correct on items to ensure that students who studied differential equations did not have a significant advantage over those who did not.

To conduct discrimination analysis, we calculated the point-biserial correlation (*r*PBIS) which reflects the extent to which higher ability students are more likely than lower ability students to select the keyed option. One item from each form were negatively correlated with the total score. One item from Purple and two from Orange had $0 < r\text{PBIS} < 0.20$. When the *r*PBIS is positive but small, it does not discriminate sufficiently among higher- and lower-scoring examinees to contribute to the overall quality of the assessment (DiBattista and Kurzawa 2011). These three items with $0 < r\text{PBIS} < 0.20$ were flagged as low-discrimination items.

Finally, we estimated the reliability of the two forms. We report Revelle's Omage Total (ω_T) as a measure of internal consistency, which is appropriate in cases where the following assumptions can be made (see Raykov 1997; Revelle and Zinbarg 2009):

1. multiple dimensions contribute to predicting the construct of interest
2. individual items measure the latent construct on different scales
3. individual items measure the latent construct with potentially differing degrees of precision

The MCQ items meet these underlying assumptions. First, mathematical modeling competence is widely presumed to be multidimensional; it draws on mathematical knowledge, real world knowledge, school-based knowledge, and knowledge of how to combine all these elements (Stillman 2000). However, when mathematical modeling is viewed in terms of its component competencies – such as the modeling cycle view adopted here – previous empirical research has indicated the interrelation and dependence of competencies rather than dimensions that could be distilled (e.g., Czocher 2018; Hankeln 2020). For example, one's capacity to validate in a given context is related to one's capacity to imagine and articulate constraints. Since all of the competencies are related to some extent, there should be a single latent variable common to all of the items contributing to a general factor (see Revelle and Zinbarg 2009).However, the individual items are on different scales aligned with the separate modeling competencies and individual items will have varying levels of error, due to the range of mathematical and real world topics they treat. Further, the items are not continuous with a normal distribution. Thus, we calculated ω_T to estimate reliability as it matches the assumptions and theoretical properties of the MCQ's development.[1]

Using the 'userfriendlyscience' package and ScaleDiagnosis() call in R, we found the ω_T values for both forms. For the Purple form, $\omega_T = 0.67$ and for the Orange form, $\omega_T = 0.63$. The scales are approaching an acceptable estimate of reliability, 0.7 with

[1] In accordance with standard practice, we provide Cronbach's α for the Purple form, $\alpha = 00.489$ and Orange form, α =0 .477. Because the MCQ instrument does not meet underlying assumptions, these are likely to be gross underestimates. See (see McNeish 2018; Peters 2014; Revelle and Zinbarg 2009) for further discussion of this issue.

highly difficult problems ($p < 0.20$) and $\omega_T = 0.71$, even accounting for the highly difficult items and the two items with negative rPBIs.

Study of the Relationship between Modeling Self-Efficacy and Modeling Competency

One approach to generating evidence for the meaningfulness of the instruments is to check whether they imply an association between self-efficacy and competency within the modeling domain. While the nature of this relationship constitutes an open question in mathematical modeling, it has been well-established for mathematics more generally. For this reason, we take the relationship as a status-quo to test the instrumentation against. We share a statistical exploration of the MCQ and MSE assessments administered prior to participation in an annual international modeling challenge (described in the next section) in order to eliminate the confounding issue of whether individuals may have gained on either measure as a consequence of participation.

Setting

We used the data collected with participants in an international modeling competition to answer our research questions. The non-profit organization SIMIODE (Systemic Initiative for Modeling Investigations and Opportunities with Differential Equations) hosts an annual modeling competition (SIMIODE Challenge Using Differential Equations Modeling [SCUDEM]) with a focus on differential equations. Since the MCQ instrument is intended to be used with STEM students who are beginning advanced studies in their majors, the sample is ideal for validation. The respondents for this part of the study are a subset of those who completed Round 3 MCQ field testing and Round 4 MSE field testing.

The competition offers participants a selection of challenging, real-world problems that they solve over the course of a week in teams of three. Each team has a faculty coach from their home institution. Each team submits a 2-page executive summary describing their solution to the problem. At the end of the week, teams convene at a local site where faculty coaches are empaneled to judge the executive summaries. The panel communicates strengths and weaknesses of the model to its team and the teams have an opportunity to revise their models, presenting final versions to other competitors at the end of the day. The panel then ranks the competitors' final models. A sample problem from the 2019 competition is given in Supplementary 2.

A total of 610 participants registered for the event. They were given a pre-event questionnaire via Survey Monkey. The pre-event questionnaire consisted of demographic questions, the MSE, one form of the MCQ, and some questions to solicit feedback about the students' experiences doing the challenge. Here we consider only the MSE, the MCQ and relevant demographic variables. A total of $n = 226$, from 119 different institutions, returned the MSE of whom 146 completed the MCQ (70 on Purple and 76 on Orange), Table 3 provides demographic information regarding the participants' genders, majors (primary major mapped to the S, T, E, and M of STEM), typical mathematics grades, and whether or not

Table 3 Participant demographic information

Major (Type)	N	Grade	N	Taken Diff Eq?	N
Science	18	A	104	Yes	31
Technology	5	B	35	No	115
Engineering	32	C	4		
Mathematics	81	D	2	Gender	N
Other	3	F	1	Male	95
None (High School)	4			Female	48

they had previously taken a course on differential equations. The "other" major category included business and economics.

Establishing a Linear Relationship

We formed the scales MSE, MCQ_Purple and MCQ_Orange summing the total score for the modeling self-efficacy assessment (out of 800) and the two versions of the competency assessment (each out of 14). We left out one item from each of the Purple and Orange forms due to a platform rendering error. Since the two forms of the MCQ were statistically equivalent in terms of overall difficulty, we formed the scale MCQ, which assigned to each participant their respective score on either MCQ_Purple or MCQ_Orange. We used the MSE and MCQ scales to address the research questions.

Figure 3 shows a scatterplot of z-scores for MCQ and MSE along with a regression line clustered by engineering or other major. It hints at a positive association between modeling self-efficacy and modeling competency as measured by the MSE and MCQ instruments. In order to confirm this relationship, we explored whether a linear relationship would appropriately model the data. A normal P-P plot of standardized residuals revealed the data points stayed close to the normal line and therefore suggested an approximately normal distribution of the errors.

The scatterplot of standardized residuals revealed a linear pattern in the data and that the variance was reasonably homogenous. Due to a perceptible degree of coning, we verified homoscedasticity according to the steps outlined in Darlington (1990) and checked whether the expected variance of studentized residues are identical for all regressor values. A univariate analysis of variance showed that there was not a significant relationship between the regressor values and the squared studentized residuals ($F(1{,}144) = 2.293$ $p = 0.132$). Therefore, the data were sufficiently homoscedastic. With these assumptions met, we were able to assume a linear relationship for the two constructs.

We conjectured we might need to consider nested nature of the data, since participants attending the same institution may have performed similarly on our measures. To explore this hypothesis, we fitted the data to a 2-level Hierarchical Linear Model (HLM) (participants within institutions) and compared this model to one where students were not nested. A log-likelihood ratio test illustrated that the nested model was

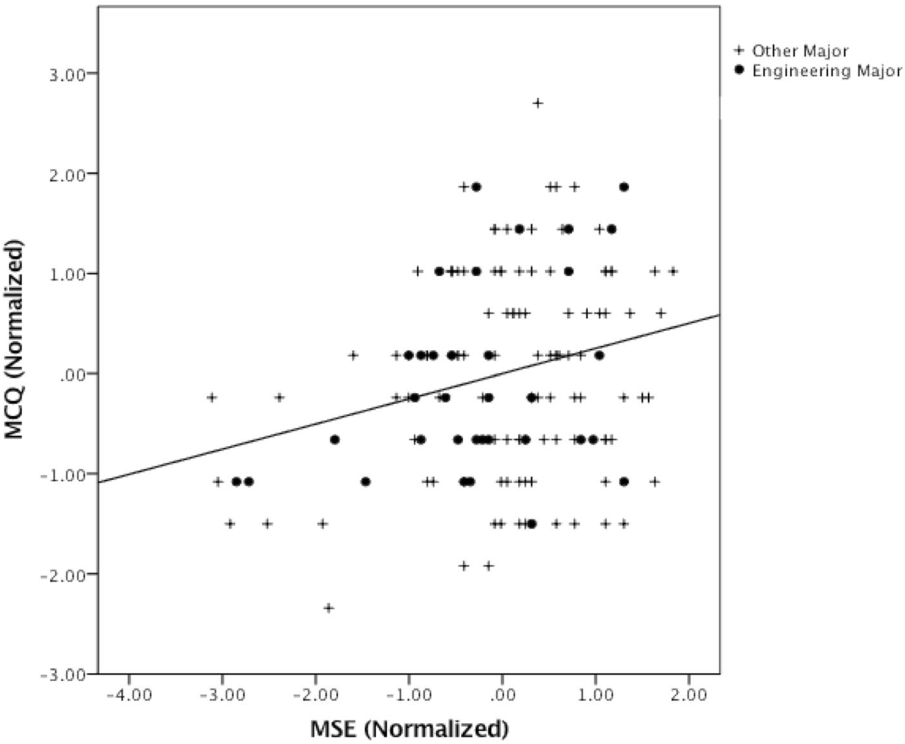

Fig. 3 Scatterplot of the normalized MCQ and MSE scores, with regression line, made in SPSS

not significantly different from the non-nested model ($\chi^2(1) = 0.127$, $p = 0.614$). We conclude that differences in performance on the MCQ across institutions were not significant. Therefore, the relationship between MCQ and MSE was best captured by a simple linear model.

Developing and Interpreting the Linear Model

In order to establish the significance and strength of the linear relationship between MSE and MCQ, we developed a series of linear models accounting for MSE and other variables that have previously been posited as relevant to modeling competency or self-efficacy. We estimated linear models that contained gender and whether the individual had taken differential equations because our previous work indicated these variables were worth considering (Czocher et al. 2019). We also considered whether the participants had competed in a modeling competition before, as a proxy for their modeling experience. Additionally, we considered STEM major as a factor and typical grade earned in previous mathematics coursework as a covariate. For each variable, we considered just a main effect model (containing MSE and the variable of interest) and interaction models (containing MSE, the variable of interest, and interaction among them). The series of models (detailed in Supplemenatry 3) provided evidence of an unanticipated result: none of the conjectured variables or their interactions were significant except for the MSE score. See Supplemenatry 3 for an overview of the coefficients from the models we tested along with corresponding p-values.

Since MSE was the only significant variable, we calculated a simple linear regression model to predict MCQ from MSE:

$$\mathrm{MCQ} = 0.252 \cdot \mathrm{MSE} \tag{1}$$

The MSE coefficient[2] (0.252) was highly significant ($t(144) = 3.130, p = 0.002$). Since we used normalized scores, the coefficient also provides information about effect size. An increase of one standard deviation on MSE predicts an increase of 0.252 standard deviations on the MCQ. When restricting our analysis to just engineering majors, counting those with a first major of engineering and those whose second major is engineering ($n = 36$) the correlation between modeling self-efficacy and competency remained significant with a larger effect size, $r(33) = 0.366,\ p = 0.033$.

Discussion and Conclusions

There is an established need within the mathematics education research community and among engineering educators to have access to well-developed measures to support program evaluation, effectiveness of instructional innovations, and to provide important necessary information about trends in students' modeling self-efficacy and competency. Within this context, this paper serves to introduce two measures that can meet both researchers' and programmatic needs. In particular, engineering programs are held to high standards and frequently expected to collect meaningful student data (Spurlin et al. 2008) and leverage the data for continual improvement. Since modeling is an essential skill for this population (Zawojewski et al. 2008), we contend that instruments like the MSE and MCQ could be useful and insightful for these stakeholders to learn about what their student populations gain from their educational experiences and make informed decisions about programmatic changes based on what the stakeholders learn.

In this paper, we shared two research-developed modeling assessments along with their known properties. We have also argued for their validity and reliability due to their grounding in empirically-derived educational theories of mathematical modeling and due to their field testing with the target populations. In doing so, we hope that these instruments can serve to mitigate some of the known impediments engineering faculty, program directors, and even researchers face when attempting to assess modeling-related educational innovations (such as trust in item and construct validity, Hoey and Nault 2008).

[2] We ran models with and without an intercept. The intercept was neither significantly nor meaningfully different from zero.

Limitations and Scope of Use

Though the sample sizes for field testing have been small, the samples themselves were repeatedly drawn from the target population. We do note that the SCUDEM participants self-selected into an extra-curricular competition, typically experienced academic success in their mathematics coursework, and therefore the engineering students participating may not be representative of engineering majors in general. Thus, some features of our sample, such as their overall academic self-efficacy and mathematical knowledge, may be higher than in the general population. However, the sample means on the MSE (prior to the annual competition) were 400.4 ($SD = 91.5$), 455.4 ($SD = 135.4$), 455.7 ($SD = 125$), 506.1 ($SD = 164.6$), indicating that even for this sub-population of highly motivated, successful, and energetic STEM majors, the MSE provided room for growth. Likewise, the MCQ items selected for future forms will be based, in part, on the empirical level of difficulty for these samples.

There are some theoretical-methodological aspects of the assessments worth discussing that may impact their appropriateness for use as an evaluation tool. First, the MSE is based in theories of self-efficacy and motivation that were largely developed with Western, specifically American, cultural values. We are optimistic about the instrument's broader use because the samples of STEM majors we have worked included individuals from international institutions, institutions with sizeable proportions of international students, and 2-year colleges. However, we would recommend that stakeholders intending to make measurements of modeling self-efficacy consider whether the construct as operationalized would be appropriate for use with the cultural groups within their inquiry setting. Second, both the MCQ and the MSE adopt an atomistic approach to measuring modeling competence and self-efficacy, focusing on individual competencies, rather than mathematical modeling as an integral skill (see Blomhöj and Jensen 2003) for further discussion of the strengths and weaknesses of each approach). While theory predicts that these competencies are interrelated and contribute towards modeling competency as a whole, the field does not yet have evidence for these claims. Thus, we would recommend complementing these assessments with rubrics developed for complex, open modeling problems. Third, from a measurement perspective it is not clear that the MCQ scale is unidimensional because its items range over modeling scenario, intended competency, and the mathematics required for each item.

Additionally, we see these instruments as providing a mechanism for researchers interested in (a) comparing instructional innovations; (b) tracking students' growth in modeling self-efficacy and competency; (c) providing a complementary individual, quantitative measure of modeling to compliment qualitative studies conducted in classrooms. We stress that the intent of the MCQ and MSE is to facilitate program and pedagogical innovations by providing feedback to the innovators, not to evaluate knowledge of individual students. Thus, we would advise exceptional caution using it to evaluate individual students, for example, when assigning grades or making admissions determinations.

Plans and Directions for Future Research

Despite our progress in developing an assessment of modeling competencies for STEM undergraduates, there remains work to increase its utility and appropriateness for its intended applications. We will continue testing more items, increasing our pool of usable items in service of developing parallel forms that are capable of measuring differences or gains in (quasi-) experimental studies whether that be comparing two (or more) experimental conditions or measuring pre−/post- gains. Our next steps will be to shift towards Item Response Theory (IRT) models to estimate assessment reliability, to create equivalent forms, to supply item-level details, and to investigate the possibility of competence sub-scales. Some promising work using IRT to investigate cohesiveness of subscales has recently begun for geometry-based modeling for secondary students (Hankeln et al. 2019) while others have investigated the feasibility of using cognitive diagnostic modeling to explore the multidimensionality of modeling assessments intended for collegiate mathematics (Alagoz and Ekici 2020). Continuing work in these directions will allow the scope of applicability of the measures to expand with increased and more varied populations. Furthermore, new samples will allow us to replicate and test the generalizability of the relationship we unearthed between MSE and MCQ.

Conclusion

Our contribution in this paper was three-fold: (1) introduce the MSE (2) introduce the MCQ (3) document the relationship between self-efficacy and competency in modeling as measured by these instruments. We have argued that the respective instruments provide a grounded means for researchers and educators to measure two important individual-level traits within the modeling domain: self-efficacy and competency. The instruments target the STEM population as they are in the midst of a crucial transition from calculus to advanced mathematics, mirroring their increasing responsibility for using mathematics to appropriately apply mathematics within their academic majors. Thus, these instruments are likely to be of value to engineering educators whose students often populate these courses and whose disciplines value mathematical modeling as a learning outcome.

Furthermore, we provided evidence that the relationship between modeling self-efficacy and modeling competency is statistically significant. Surprisingly, we found self-efficacy to be the only significant predictor of competency. This analysis allows us to make two claims: (1) there is a linear relationship between modeling self-efficacy and competency and (2) this relationship does not account for all variation in competency (i.e., the coefficient was 0.252 standard deviations overall and 0.366 for engineering students). We conclude that both instruments measure important, related, but ultimately independent constructs critical to modeling. Thus, plans for assessing modeling as a learning outcome should consider measuring both the self-efficacy and competence constructs because both can be vital for students' persistence in engineering majors.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40753-020-00124-7.

Funding This material is based upon work supported by the National Science Foundation under Grant No. 1750813.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdiciplinaty reviews: computational statistics, 2*, 433–459.

ABET Engineering accreditation Comission (2018-2019). Criteria for accrediting engineering programs.

Alagoz, C., & Ekici, C. (2020). Cognitive diagnostic modelling for mathematical modelling assessment. In G. Stillman, G. Kaiser, & C. Lampen (Eds.), *Mathematical modelling education and sense-making* (pp. 349–359). Cham: Springer.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (Vol. 3, pp. 307–337). United States of America: Information Age Publishing.

Betz, N., & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational Behavior, 23*(3), 329–345.

Bliss, K., Libertini, J., Levy, R., Zbiek, R. M., Galluzzo, B., Long, M., et al. (2016). *Guidelines for assessment and instruction in mathematical modeling education*. USA: Consortium for Mathematics and Its Applications, Society for Industrial and Applied Mathematics.

Blomhöj, M., & Jensen, T. (2003). Developing mathematical modelling competence: Conceptual clarification and educational planning. *Teaching Mathematics and Its Applications, 22*(3), 123–139.

Blum, W., & Leiss, D. (2007). How do students and teachers deal with modelling problems? In C. Haines, P. L. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical modelling (ICTMA 12): Education, engineering and economics* (pp. 222–231). Chickester: Horwood Publishing Limited.

Bressoud, D., & Rasmussen, C. (2015). Seven characterists of successful calculus programs. *Notices of the American Mathematical Society, 62*(2), 144–146.

Chiel, H. J., McManus, J. M., & Shaw, K. M. (2017). From biology to mathematical models and back: Teaching modeling to biology students, and biology to math and engineering students. *CBE Life Sciences Education, 9*(3), 141–377.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Czocher, J. (2016). Introducing modeling transition diagrams as a tool to connect mathematical modeling to mathematical thinking. *Mathematical Thinking and Learning, 18*(2), 77–106.

Czocher, J. (2017). How can emphasizing mathematical modeling principles benefit students in a traditionally taught differential equations coures? *The Journal of Mathematics Behavior, 45*, 78–94.

Czocher, J. (2018). How does validating activity contribute to the modeling process? *Educational Studies in Mathematics, 99*, 137–159.

Czocher, J. (2019). Precision, priorities, and proxies in mathematical modelling. In G. Stillman & J. Brown (Eds.), *Lines of inquiry in mathematical modelling research in education* (pp. 105–123). New York: Springer.

Czocher, J. A., & Kandasamy, S. S. (2018). On how participation in a modeling competition occasions changes in undergraduate students' self-efficacy regarding mathematical modeling. In T. E. Hodges, G. J. Roy, & A. M. Tyminski (Eds.), *Proceedings of the 40th annual meeting of the north American chapter of the international group for the psychology of mathematics education*. Greenville, SC: University of South Carolina & Clemson University.

Czocher, J., Melhuish, K., & Kandasamy, S. S. (2019). Building mathematics self-efficacy of STEM undergraduates through mathematical modeling. *International Journal of Mathematical Education in Science and Technology*. https://doi.org/10.1080/0020739X.2019.1634223.

Darlington, R. (1990). *Regression and linear models*. New York: McGraw-Will.

DiBattista, D., & Kurzawa, L. (2011). Examination of the quality of multiple-choice items on classroom texts. *The Canadian Journal for the Scholarship of Teaching and Learning, 2*(2), 1–23.

Dominguez-Garcia, S., Garcia-Planas, M. I., & Taberna, J. (2016). Mathematical modelling in engineering: An alternative way to teach linear algebra. *International Journal of Mathematical Education in Science and Technology, 47*(7), 1076–1086.

Dym, C. L. (2004). *Principles of mathematical modeling*. Burlingon, MA: Elsevier Inc..

Eccles, J. S., & Wang, M.-T. (2016). What motivates females and males to pursue careers in mathematics and science? *International Journal of Behavioral Development, 40*(2), 100–106.

European Network for Accreditation of Engineering Education (2018). EUR-ACE framework standards and guidlines. https://www.enaee.eu/eur-ace-system/standards-and-guidelines/#standards-and-guidelines-for-accreditation-of-engineering-programmes. Accessed 6 June 2020.

Fauconnier, G. (2001). Conceptual blending. *The encyclopedia of the social and behavioral sciences,190,* 2495–2498. https://doi.org/10.1016/B0-08-043076-7/00363-6.

Frejd, P. (2013). Modes of modelling assessment—A literature review. *Educational Studies in Mathematics, 84*(3), 413–438.

Geisinger, B. N., & Rajraman, D. (2013). Why they leave: Understanding student attrition from engineering majors. *International Journal of Engineering Education, 29*(4), 914–925.

Güner, N. (2013). Senior engineering students' views on mathematics courses in engineering. *College Student Journal, 47*(3).

Hackett, G., & Betz, N. (1989). An exploration of the mathematics self-efficacy/mathematics performance correspondence. *Journal for Research in Mathematics Education, 20*(3), 261–273.

Haines, C., Crouch, R. & Davis, J. (2000). Mathematical Mmdelling skills: A research instrument. Technical Report No. 55. University of Hertfordshire Faculty of Engineering and Information Sciences.

Hallström, J., & Schönborn, K.,. J. (2019). Models and modelling for authentic STEM education: Reinforcing the argument. *International Journal of STEM Education, 6*(22). Available from: https://stemeducationjournal.springeropen.com/articles/10.1186/s40594-019-0178-z.

Hankeln, C. (2020). Mathematical modeling in Germany and France: A comparison of students' modeling processes. *Educational Studies in Mathematics, 103*, 209–229.

Hankeln, C., Adamek, C., & Greefrath, G. (2019). Assessing sub-competencies of mathematical modelling-development of a new test instrument. In G. Stillman & J. Brown (Eds.), *Lines of inquiry in mathematical modelling research in education (ICME-13 monographs)*. Springer international publishing. Available from: https://link.springer.com/chapter/10.1007%2F978-3-030-14931-4_8.

Hoey, J. J., & Nault, E. W. (2008). Barriers and challenges to assessment in engineering education. In J. P. Lavel, S. A. Rajala, & J. E. Spurlin (Eds.), *Designing better engineering education through assessment: A pratical resourse for faculty and development chairs on using assessment and ABET criteria to improve student learning* (pp. 171–189). Sterling, VA: Stylus Publishing.

International Engineering Alliance (2014). 25 years of the Washington accords. (pp. 1–25). Accessed from: https://www.ieagreements.org/assets/Uploads/Documents/History/25YearsWashingtonAccord-A5booklet-FINAL.pdf.

Jones, B. D., Paretti, M. C., Hein, S. F., & Knott, T. W. (2010). An analysis of motivation constructs with first-year engineering students: Relationships among expectancies, values, achievement, and career plans. *Journal of Engineering Education, 99*, 319–336.

Kaiser, G. (2017). The teaching and learning of mathematical modeling. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 627–291). United States: The National Council of Teachers of Mathematics, Inc.

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education, 3*(11), 1–11.

Kertil, M., & Gurel, C. (2016). Mathematical modeling: A bridge to STEM education. *International Journal of Education in Mathematics, Science and Technology, 4*(1), 44–55.

Kim, S.-I., Jiang, Y., & Song, J. (2015). The effects of interest and utility value on mathematics engagement and achievement. In K. A. Renninger, M. Nieswandt, & S. Hidi (Eds.), *Interest in mathematics and science learning* (pp. 63–78). Washington, DC: American Educational Research Association.

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..

Liu, H., & Raghavan, J. (2009). A mathematical modeling module with system engineering approach for teaching undergraduate students to conquer complexity. In G. Allen, J. Nabrzyski, E. Seidel, G. D. van Albada, J. Dongarra, & P. M. A. Sloot (Eds.), *Computational sciences - ICCS 2009 (pp. 93–102)*. Berlin Heidelberg: Springer.

Loo, C. W., & Choy, J. L. F. (2013). Sources of self-efficacy influencing academic performance of engineering students. *American Journal of Educational Research, 1*(3), 86–92.

Maaß, K. (2006). What are modelling competencies? *ZDM: Mathematics Education, 38*(2), 113–142.

Malmqvist, J., Edstrom, K., Gunnarsson, S., & Ostlund, S. (2006). The application of CDIO standards in the evaluation of Swedish engineering degree programmes. *World Transactions on Engineering and Technology Education, 5*(2), 361–364.

McGourty, J., Sebastian, C., & William, S. (2013). Developing a comprehensive assessment program for engineering education. *Journal of Engineering Education, 87*(4), 355–361.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412–433.

Moore, T. J., Miller, R. L., Lesh, R., Stohlmann, M. S., & Kim, Y. R. (2013). Modeling in engineering: The role of representational fluency in students' conceptual understanding. *Journal of Engineering Education, 102*(1), 141–178.

Niss, M., Blum, W., & Galbraith, P. L. (2007). Introduction. In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education* (pp. 2–32). New York, NY: Springer.

Pennel, S. (2009). An engineering-oriented approach to the introductory differential equations course. *Problems, Resources, and Issues in Mathematics Undergraduate Studies, 19*(1), 88–99.

Peters, G.-J. Y. (2014). The alpha and the omega of scale reliability and validity. *The European Health Psychologist, 16*(2), 56–69.

Rasmussen, C., & Kwon, O. N. (2007). An inquiry-oriented approach to undergraduate mathematics. *Journal of Mathematical Behavior, 26*, 189–194.

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research, 32*(4), 329–353.

Revelle, W., & Zinbarg, R. (2009). Coefficients alpha, beta, omega and the glb: Comments on Sijtsma. *Psychometrika, 74*(1), 145–154.

Schukajlow, S., Liess, D., Pekrun, R., Blum, W., Müller, M., & Messner, R. (2012). Teaching methods for modelling problems and students' task-specific enjoyment, value, interest and self-efficacy expectations. *Educational Studies in Mathematics, 79*, 215–237.

Sokolowski, A. (2015). The effect of math modeling on student's emerging understanding. *The IAFOR Journal of Education, 3*(3), 142–156.

Spurlin, J. E., Rajala, S. A., & Lavel, J. P. (2008). *Designing better engineering education through assessment:A practical resource for faculty and department chairs on using assessment and ABET criteria to improve student learning*. Sterling, Virginia: Stylus Publishing, LLC.

Stillman, G. (2000). Impact of prior knowledge of task context on approaches to applications tasks. *Journal of Mathematical Behavior, 19*, 333–361.

Stillman, G., & Galbraith, P. L. (1998). Applying mathematics with real world connections: Metacognitive characteristics of secondary students. *Educational Studies in Mathematics, 36*(2), 157–194.

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and thinks, women and people: A meta-analysis of sex differences in interest. *Psychological Bulletin, 1135*(6), 859–884.

Temponi, C. (2005). Continuous improvement framework: Implications for academia. *Quality Assurance in Education, 13*(1), 17–36.

Trigueros, M., & Possani, E. (2013). Using an economics model for teaching linear algebra. *Linear Algebra and its Applications, 438*(4), 1779–1792.

Yildirim, T., Shuman, L. J., & Besterfield-Sacre, M. (2010). Model-eliciting activities: Assessing engineering student problem solving and skill integration processes. *International Journal of Engineering Education, 26*(4), 831–845.

Young, C. Y., Georgiopoulos, M., Hagen, S. C., Geiger, C. L., Dagley-Falls, M. A., Islas, A. L., et al. (2011). Improving student learning in calculus through applications. *International Journal of Mathematical Education in Science and Technology, 42*(5), 591–604.

Zawojewski, J. S., Hjalmarson, M. A., Bowman, K. J., & Lesh, R. (2008). A modeling perspective on learning and teaching in engineering education. In J. S. Zawojewski, H. A. Diefes-Dux, & K. J. Bowman (Eds.), *Models and modeling in engineering education*. Leiden, The Netherlands: Brill.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.