Content and internal structural validity of the motivation for mathematics abbreviated instrument

Kenneth Butler¹ · Robert F. Dedrick²

Received: 11 August 2020 / Accepted: 14 October 2021 / Published online: 15 November 2021
© The Author(s) 2021

Abstract

This study details the development and initial validation of a 16-item psychometric instrument measuring college students' motivation for mathematics. The Motivation for Mathematics Abbreviated Instrument (MMAI) consists of four factors (intrinsic motivation, mastery orientation, performance orientation, and expectancy) that were derived from three predominant motivation theories: self-determination, achievement goal, and expectancy value. The instrument development process began with an initial pool of 107 items from existing motivational measures. Input from members of the *American Education Research Association (AERA) Special Interest Group (SIG) for Motivation in Education*, interviews with students and instructors, and factor analyses of a preliminary ($n = 186$ post-secondary students) and final administration ($n = 386$ post-secondary students) of the instrument were used to create the final version of the 16-item measure. Qualitative analyses verified content. Results of the confirmatory factor analyses indicated acceptable fit of the 16-item four-factor model and measurement invariance by gender. We found strong composite reliability. Average variance extracted revealed acceptable internal validity. We conclude that the MMAI is a brief, psychometrically sound measure of motivation that provides a practical and useful tool for classroom-based research into motivation for mathematics.

Keywords Motivation · Mathematics education · Survey · Factor analysis · Invariance · Mixed methods

✉ Kenneth Butler
kbutler2@udayton.edu

Robert F. Dedrick
dedrick@usf.edu

¹ Mathematics Education, University of Dayton, Dayton, OH, USA

² Educational Measurement, University of South Florida, Tampa, FL, USA

Introduction

Theory and measurement are inextricably linked (Cronbach and Meehl 1955). Well-developed theories are crucial to the development of sound measurement instruments, and sound measurement plays an important role in advancing theory. The link between measurement and theory is particularly evident in the realm of motivation. For motivation, there are a number of theories, including self-determination (Ryan and Deci 2000), achievement goals (Elliot and Harackiewicz 1996), and expectancy value (Eccles 1987). These theories have guided the development of instruments that include the *Intrinsic Motivation Inventory* (McAuley et al. 1987), the *Patterns of Adaptive Learning Scale (PALS)* (Midgley et al. 2000), and the *Motivated Strategies for Learning Questionnaire (MSLQ)* (Pintrich et al. 1992). These instruments have facilitated research on motivation that has resulted in extensions, elaborations, and refinements of motivation theory (Hancock 2007; Li 2017; Monteiro et al. 2015; Simper et al. 2016; Wibrowski et al. 2017; Yin and Wang 2016).

Although the previously listed motivation instruments have a number of strengths, one of their limitations is their length. For example, the PALS provides a 94-item survey, and the MSLQ uses 31 items to measure motivation and another 50 to measure learning strategies. The number of items in these surveys can make them impractical to use in ecologically valid research (Gogol et al. 2014). Concerns about the practicality of measures with a large number of items have also been raised by Kosovich et al. (2017), who propose that long questionnaires may produce negative reactions in the respondent and a refusal to complete the items. These problems are exacerbated in situations in which many constructs are to be measured, and when measures are administered frequently to monitor constructs over time. Additional calls for shorter measures of psychological constructs have been made by Coelho et al. (2018) and others (e.g. Rammstedt and Beierlein 2014). The development of a valid abbreviated instrument to measure motivation is important because researchers have expressed a need for short, quickly administered, psychometrically sound instruments for research aligned with theory. Although there are many current instruments that are psychometrically sound and widely used, none were designed to discriminate between motivation constructs with this focus on brevity. As the *Motivation for Mathematics Abbreviated Instrument (MMAI)*—the focus of this paper—was developed to be brief, psychometrically sound, and aligned with theory, it fills this niche. The resulting 16-item instrument uses four items to measure each of four constructs: intrinsic motivation, mastery orientation, performance orientation, and expectancy. The current study focuses on the development of this instrument and how this instrument performs with undergraduate students in general education mathematics courses.

In the following sections, we describe the motivational theories that provide the conceptual framework for the MMAI and then present the procedures for constructing the instrument. Afterwards, we discuss evidence from multiple sources used to verify content for our population and support an argument for the internal structural validity of the MMAI. We used the *Standards for Educational and*

Psychological Testing (American Educational Research Association et al. 2014) to frame the validity arguments.

Theoretical framework

The MMAI measures four motivational constructs and is based on three predominant theories of motivation: self-determination theory (Ryan and Deci 2000), achievement goal theory (Elliot and Harackiewicz 1996), and expectancy value (Eccles 1987).

Self-determination

This theory is a cognitive approach to motivation, as it relies on innate inner workings of human cognition in understanding motivation. Here, inner processes actively promote actions instead of actions being passive responses to stimuli (Deci and Ryan 1985; Ryan et al. 1997; Ryan and Deci 2000). According to self-determination theory, motivation is a continuum from intrinsic motivation through various levels of extrinsic motivation to amotivation. These motivational classifications are characterized by the level of autonomy and the source of regulation for one's engagement. When people are completely autonomous in their actions, engage in an activity for pure enjoyment, interest, or inherent satisfaction and are self-regulated, they are understood to be intrinsically motivated. On the other end of the spectrum, when people have no autonomy, are forced to engage in an activity, and are regulated by others, they are understood to be amotivated. Various levels of extrinsic motivations lie between these two extremes. Students are intrinsically motivated in mathematics if they autonomously engage in mathematics for enjoyment and interest in the subject.

Achievement goals

This theory (Elliot and Harackiewicz 1996) provides a rationale as to why one chooses to learn. Motivations for mastery-oriented people are rooted in learning for understanding. In the context of mathematics, students have mastery orientation towards mathematics if they engage with content because they want to understand mathematics. This is in contrast to performance orientation. The motivations for performance-oriented people are rooted in a desire to be compared favourably with others. Students may be motivated to approach a task based on mastery or performance orientation, but they may also avoid a task based on mastery or performance orientation. Students may avoid mastering a subject because they believe this may threaten their personal belief system, or they may avoid taking a test or doing homework because they do not want to be compared unfavourably with others. This approach reveals a four-factor model with orientations for mastery and performance approach, or mastery and performance avoidance.

Expectancy value

This theory grew out of behaviourism, but it has expanded to include cognitive dimensions tying expectations of success to the value associated with succeeding (Eccles 1987; Eccles and Wigfield 1995). In this theory, individuals will be motivated to perform an action if they believe they can perform it successfully, and if they believe there is value in success. The expectation of success is intimately linked to one's perceived efficacy within the domain, and to the value placed on being successful (Eccles 1987). In mathematics classrooms, expectancy is modified by students' perceived abilities to understand content, and their faith that they will receive a positive grade. Closely connected with this, task value is a multi-faceted construct associated with the perceived benefits and costs of succeeding. In this framework, students need to believe they can be successful, and they need to see value in their effort.

Research questions

This study details the development and evaluation of the MMAI when administered to post-secondary mathematics students. The focus, therefore, is on the psychometric performance of the items and the instrument. Two research questions relate to measurement validity that focus on evidence of content and internal structural validity:

1. To what extent do the items in the MMAI represent their intended constructs for undergraduate general education math students?
2. To what extent does a latent factor measurement model consisting of four factors (intrinsic motivation, mastery orientation, performance orientation, and expectancy) underlying the MMAI fit the response data for our targeted population?

Methods

Research design

The focus of this study is on the development and evaluation of the psychometric quality of items measuring four motivational factors (intrinsic motivation, mastery orientation, performance orientation, and expectancy). As the focus of this study was survey items, it was considered exempt by an Institutional Review Board. Informed consent was provided to all participants explaining that their participation was voluntary and no personal identifying information would be kept. All participants were over 18 years old. We conducted this study in five phases. See Table 1 for a summary. The first phase involved the compilation of items from popular existing instruments to represent motivational constructs. We strove to keep the roots and

Table 1 Summary of steps used to develop the motivation for mathematics abbreviated instrument (MMAI)

Phase	Tasks/analyses	Number of items
1	Create an initial items pool from existing measurement instruments grounded in different motivational theories (Self-determination, Achievement goal, and Expectancy value)	107
2	Use input from 123 members of the American Education Research Association (AERA) Special Interest Group for Motivation in Education to identify a reduced number of items measuring seven motivational constructs (intrinsic motivation, extrinsic motivation, mastery orientation, performance orientation, self-efficacy, task value, or expectancy)	41
3	Use cognitive interviews with 4 post-secondary students (2 males and 2 females; 2 students who like math and 2 students who did not) to identify items that best represent the intended constructs	26
4	Use IRT and Exploratory Factor analyses of response data from 186 post-secondary students to identify a theoretically and statistically meaningful reduced set of items to measure the intended constructs	16
5	Conduct confirmatory factor analyses of response data from 386 post-secondary students to evaluate model fit of the resulting Motivation for Mathematics Abbreviated Instrument (MMAI)	16

meaning of the items the same and only change their focus towards mathematics. Each of the 107 items representing intrinsic motivation, extrinsic motivation, mastery orientation, performance orientation, self-efficacy, expectancy, or task value were sourced from popular existing instruments. The fact that we did not change the overall content of the items, which have been previously vetted in multiple studies, provides support for content validity of the items. Informed consent was provided to all participants explaining that their participation was voluntary and no personal identifying information would be kept. All participants were over 18 years old.

For the second phase, we had the members of the AERA SIG for Motivation in Education provide input on the items using an online survey (e.g. select the 5 best of many good items that represent expectancy). By having educational researchers who focus on motivation review the items, we verified a general agreement that the items moving to the third phase are representative of their intended constructs. This step was designed to enhance the content validity of the items. Based on this analysis, we removed items that had less than 40% endorsement, and trimmed the initial pool of 107 items down to 41 items.

For the third phase, our goal was to reduce the number of items so that the resulting instrument could be administered quickly, would be representative of the intended constructs, and would be able to discriminate between constructs. We used both qualitative and quantitative sources of data to achieve this goal. We began with stratified sampling to select four students and two instructors for cognitive interviews to identify items that best aligned with the motivational theories guiding instrument development. Tourangeau (1984) provided our interview framework, and Straus and Corbin (1990) provided the constant comparison methodology for coding transcripts. We then quantified this qualitative data using intra-respondent matrices (Onwuegbuzie and Teddlie 2003). Through these processes, we reduced the number of items from 41 to 26.

For the fourth phase, we administered this preliminary 26-item instrument to 186 undergraduate math students, and we analysed the data using two separate analyses. *Mplus* 8.4 (Muthén and Muthén 1998–2017) was used to conduct all analyses. First, we conducted an IRT analysis using the logit link function. To determine if there were significant differences in item loadings, an IRT model with item loadings constrained was compared to model where the item loadings were allowed to vary. If the model, where item loadings were allowed to vary, has a significantly better fit, then we can surmise the item loadings were different. As the items were polytomous with a response scale of 1 to 5 where 1 = *Not at all true* and 5 = *Very true*, we utilized a Graded Response Model (Samejima 1996). Assuming the item loadings were different, we then examined the item response curves and item information curves to determine how the items were behaving at the individual and at the instrument level. Items were considered to perform poorly if their item response curves did not show discrimination between answers, or if their information curves were flat containing little area.

Following Costello and Osborne's (2005) framework, we examined this same data utilizing EFA with the goal of understanding the underlying factor structure of the data. We assumed the data to be continuous, and used maximum likelihood robust (MLR) estimation. Although this method uses different computations

for standard errors, final parameter estimates are the same as they would be under Maximum Likelihood estimation. We used Kaiser's criterion, Parallel analysis, and visual analysis to determine the number of factors to extract. As all intended constructs in this study measure motivation, the factors should be correlated, and so we employed Geomin rotation. We then examined item loadings using this rotated pattern matrix. If an item did not load on its intended construct or loaded on more than one construct, then the item was considered to be performing poorly. The cut-off for a loading was set at 0.40. By removing items that did not perform well under IRT and EFA, and because of a desire to align the instrument with theory, we reduced the number of items from 26 to 16. This resulted in the four-factor 16-item MMAI.

For the fifth and final phase, we administered the 16-item MMAI to 386 undergraduate students, and we conducted a confirmatory factor analysis (CFA) using *Mplus* 8.4 (Muthén and Muthén 1998–2017). We used Maximum Likelihood Robust extraction. The model specified was theoretically based and comprised four latent exogenous factors each with four indicators. The measurement equation is $\mathbf{x} = \mathbf{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}$, where $\mathbf{x}$ is a (16×1) vector of observed variables, $\mathbf{\Lambda}_x$ is a regression coefficient matrix of $\mathbf{x}$ on $\boldsymbol{\xi}$, $\boldsymbol{\xi}$ is a (4×1) vector of latent constructs derived from $\mathbf{x}$, and $\boldsymbol{\delta}$ is a (16×1) vector of errors in measurement in $\mathbf{x}$. As is the default in *Mplus*, we assumed the errors of measurement in $\mathbf{x}$ were uncorrelated, i.e. the covariance matrix $\boldsymbol{\delta}$ is diagonal.

To test how well our model fit the data, we relied on several commonly used fit indices and set our criteria for demonstrating a good fit at less than 0.08 for the standardized root mean square residual (SRMR), less than 0.06 for the root mean square error of approximation (RMSEA), and above 0.95 for the comparative fit index (CFI) as per work by Browne and Cudeck (1993), MacCallum et al. (1996), and Hu and Bentler (1999). Sun (2005) provides a good discussion of fit indices and applications of CFA.

To examine internal reliability and convergent validity, we compare the composite reliability and the average variance extracted (AVE) to the theoretical cut scores proposed by Fornell and Larcker (1981), and Bagozzi and Yi, (1988). Composite

reliability of a latent construct ξ_j is $\rho_{CR(\xi_j)} = \frac{(\sum_{i=1}^4 \lambda_{ji})^2}{(\sum_{i=1}^4 \lambda_{ji})^2 + \sum_{i=1}^4 \sigma_{\delta_i}^2}$, where λ_{ji} is the standardized loading for the i^{th} item associated with the factor ξ_j , and $\sigma_{\delta_i}^2$ is the variance for the measurement error associated with this same item. We use composite reliability above 0.7 as desirable. AVE of a latent construct ξ_j is $\rho_{AVE(\xi_j)} = \frac{\sum_{i=1}^4 \lambda_{ji}^2}{\sum_{i=1}^4 \lambda_{ji}^2 + \sum_{i=1}^4 \sigma_{\delta_i}^2}$.

An AVE above 0.5 reveals good convergent validity. To examine discriminant validity, we compare AVE to the squared correlations between factors. Discriminant validity is acceptable when AVE of any factor is greater than the squared correlations between it and all other factors. Thus, $\rho_{AVE(\xi)} > \gamma^2$, where γ is the correlation of the latent factor ξ with any other factor, shows good discriminant validity. We also provide individual item reliabilities as $\rho_{x_i} = \frac{\lambda_i^2 * \sigma_{\xi}^2}{\lambda_i^2 * \sigma_{\xi}^2 + \sum_{i=1}^4 \sigma_{\delta_i}^2}$, where σ_{ξ}^2 is the variance of the factor ξ associated with the x_i . Ab Hamid et al. (2017) found an item should be removed when ρ_{x_i} falls below 0.4.

We then conducted invariance testing to determine if the instrument could be used as an unbiased measure of differences between males and females based on self-reported gender. In answer to the statement on the questionnaire, “I personally identify as: male, female, other”, 167 students identified as male, and 211 identified as female. Eight selected other or did not answer. Although many grouping variables could have been used in the invariance testing (e.g. race, ethnicity), we decided to begin with gender in view of the extensive research that has examined differences in motivation for mathematics by gender.

To test measurement invariance, we began by checking the fit of the model where loadings, intercepts, and errors were allowed to vary between groups—the configural model. Using this as the base model and working through the metric, scalar, and finally the strict invariance model, we examined differences in fit between successive models. More detailed steps are provided in the results section. Good sources for information on measurement invariance are Meredith (1993), Putnick and Bornstein (2016), Schmitt and Kuljanin (2008), and Vandenberg and Lance (2000).

Results

Phase 1: item selection

The four factors targeted by the *Motivation for Mathematics Abbreviated Instrument* (MMAI) are intrinsic motivation, mastery orientation, performance orientation, and expectancy. The development of MMAI began with the creation of a large pool of items from several previously published instruments measuring disparate theories of motivation. We chose these specific instruments (listed in the next few paragraphs) because of their widespread popularity in educational research. By including items from instruments with a strong record of utility in research, we began with some inherent evidence for validity of content.

To represent self-determination theory, we selected all items representing interest and enjoyment from the *Intrinsic Motivation Inventory* (McAuley et al. 1987). This article has been cited at least 1988 times. All items representing intrinsic and extrinsic motivation from the *Sport Motivation Survey* (Pelletier et al. 1995) were included, and this article has been cited at least 1948 times. Items representing intrinsic motivation from Pintrich et al.’s (1992) *Motivated Strategies for Learning Questionnaire* were also included. This resulted in the inclusion of 34 items representing self-determination theory. To represent achievement goal theory, we selected all items representing mastery orientation and performance orientation from the *Patterns of Adaptive Learning Scale* (Midgley et al. 2000). This article has been cited at least 1527 times. This resulted in the inclusion of 11 items representing mastery orientation and 10 items representing performance orientation. To represent expectancy value, we selected all items intended to measure expectancy from the *Motivated Strategies for Learning Questionnaire* (Pintrich et al. 1992). This article has been cited at least 1339 times. We also included the items from this survey representing task value as well as items from a *Perceived Instrumentality Survey* by

Miller et al. (1999). This resulted in the inclusion of 12 items representing expectancy and 11 items representing task value.

Self-efficacy is the belief that you can be successful in a limited domain as described by Bandura (1997). It is a strong predictor of success and is closely related to motivation. To represent self-efficacy, we included the 24 items from Usher and Pajares' (2009) *Sources of Self-Efficacy* survey.

As the intent was to measure motivation for mathematics, the roots of the items were left unchanged with only slight modifications to focus the items on mathematics. For example, "Learning this material is enjoyable" from Miller et al. (1999, p. 255) was changed to "Learning math is enjoyable". Other than focussing the item on our intended domain, these changes had little impact on the content of the item.

Phase 2: expert review

After assembling the initial pool of 107 items, we sent emails to all 511 members of the AERA SIG for Motivation in Education asking them to evaluate the items in these surveys. These emails informed members that the purpose of the study was to develop and initially validate an abbreviated instrument intended to measure constructs from several of the predominant theories of motivation. The members were informed that the intended population was post-secondary students in the United States, and that the items utilized a five-point scale. We attached an informed consent form to these emails explaining the voluntary and anonymous nature of participation in this study. These emails also contained a request for any questions, comments, or concerns about the study.

We asked the members to select the five items that best represented each of the seven intended constructs. The items also needed to discriminate between constructs. We also asked for commentary. We received 123 responses (24% response rate) to the survey. We then ranked the items based on the members' endorsements and analysed the comments of the item reviewers. Based on these data, 41 items representing intrinsic motivation, extrinsic motivation, mastery orientation, performance orientation, self-efficacy, expectancy, and task value became the preliminary instrument, and all had at least 40% endorsement.

In comments concerning the intrinsic motivation items, a few SIG members highlighted difficulties in separating intrinsic motivation from mastery; however, many of the comments suggested that the variable separating these constructs was enjoyment of a task. As one member wrote, "I think intrinsic motivation is all about the joy someone gets from a task, just because it is that task", and another wrote, "intrinsic motivation speaks to something being fun and enjoyable just because". Another SIG member wrote, "I had difficulty with doing things for some pleasure as [being] directly intrinsic, but [the intrinsic motivation items] can be reduced to five with the following words: satisfying, enjoyment, pleasure, excitement and fun". One of the intrinsic motivation items in the MMAI references fun, one references enjoyment, one references excitement, and one references interest.

Many of the comments concerning the mastery orientation items discussed improving and learning. One member wrote, "Mastery orientation is defined as

improving”, and another wrote, “Mastery orientations are expressed when a person tends to set objectives of learning as much as possible”. Another member had a slightly different view of mastery. “The focus, in my opinion, should be about the task-based standards of mastery and understanding, and less so on the self-improvement standard”. Two of the mastery orientation items in the MMAI reference learning, one references understanding, and one makes specific reference to mastery.

Comments about performance orientation items tended to focus on being better than others. As one member wrote, “Here it is important to decide whether you want to capture via the performance approach (PA) goal items; (a) the aim of being better than others; (b) the aim of demonstrating one’s superiority; (c) both... I favor (a) over (b) in defining PA goals”. All the performance orientation items in the MMAI focus on performing better than others.

Regarding most of the expectancy items members stated that the items were confounded with self-efficacy. As one wrote, “I have checked the ones most like expectancy, which is actually the same thing as self-efficacy”. Thus, the commentary about the content of the included expectancy items revealed the strong relationship between expectancy and self-efficacy. This relationship is aligned with the expectancy-value framework espoused by Eccles and Wigfield (1995).

Phase 3: cognitive interviews

We used cognitive interviews (Tourangeau 1984; Willis 2004) with four students and two instructors to obtain feedback about the clarity and alignment of the 41 items with the four theoretical constructs underlying the MMAI. Unlike the theoretical nature of the feedback from experts, the cognitive interview data obtained from students and instructors were from the target population being measured.

Students were recruited from one math class at a public university in the southern United States. We began by asking students to fill out cards with their names, gender, and a response on a scale of one to ten to a question about whether they liked mathematics. From these cards, we used stratified sampling to select participants so that we had equal representation based on gender and liking mathematics. This resulted in one male who liked mathematics and one who did not—let us call them Bill and Ted, and one female who liked mathematics, and one who did not—let us call them Thelma and Louise. Two instructors were also interviewed.

Students were asked the following questions for each item:

1. What was the meaning of (insert item)?
2. What information did you need to answer (insert item)?
3. What judgements did you make when answering (insert item)?
4. What was your answer to (insert item)?

Instructors were asked similar questions; however, questions 2 and 3 were modified to ask what information or judgements a student might need to answer. These interviews provided over four hours of audio recordings.

We then transcribed these interviews, and two researchers separated these transcripts into concept phrases. For this process each comment was discussed, and when phrases within a comment could stand alone as distinct ideas, then the comment was separated into individual concept phrases. Then, we entered these concept phrases row by row into an intra-respondent matrix (Onwuegbuzie 2003). This resulted in 2764 rows of data. We then coded these concept phrases using a constant comparison methodology (Straus and Corbin 1990). In the coding process, at least 41 different codes were considered. Through a process of attrition and conglomeration, 10 codes emerged. Table 2 lists the codes and keywords used in this process.

Three of the intrinsic motivation items invoked fun and enjoyment, and two of the interviewees explicitly referenced these terms. As Bill stated, “It’s enjoyable because of the complexity”, and Ted said, “Some stuff that is challenging can be enjoyable”. These two students also commented, “It’s fun to learn new things”, and “If something is fun, you enjoy doing it”. The intrinsic motivation item describing mathematics as interesting led interviewees to reflect on engagement, mastery, and relevance. This was often stated in a negative light. Louise said, “[Mathematics] is just one way of doing something”. When questioned for clarity, she stated, “Interesting would be something where you could be creative?”. Although not an endorsement, the concept of being creative may have implied autonomy, and this implied cognitive processes that were associated with intrinsic motivation.

Some of the responses to the mastery orientation items concerned wanting to improve. Thelma states she “would want to learn new math concepts because [she] wants to progress mathematically”, or as Bill states, “It’s always good to learn new things and I would like to master some of those things”. One mastery item invoked mastery avoidance processes. As Ted states, “I wouldn’t want to have a mathematician mindset for understanding mathematics”. Whether the items invoked a desire

Table 2 Codes and keywords used in analysis

Codes	Keywords
Fun	fun, enjoy, like, excite, want to
Mastery	grasp, learn, understand (working towards understanding), complex, master, getting it, mistake
Performance	grade, right, test, compete, exam, pass, GPA, best, prove, well, percentage, compare, average, points, benchmark, measure, competition
Relevance	apply, life, every day, world, career, future, major, goal, employer, relate, value, occupation, important, reward, benchmark, win, lose, award, recognition, gratification
efficacy	mistake, difficult, get it, ability, smart, confident, understand (to possess understanding), grasp, hard, comfort, know, retain, no clue, literate
Engage	speak up, contribute, immersed, challenge, tried, push, hours, engrossed, interesting, effort, practice, work, giving up, passion, repeat, struggle, over and over, do your best
Affect	hate, feel, humble, happy, bad, relax, stress, passion, discourage, stupid, care, embarrass, pressure, confused, risk, jumbled, scare
Relatedness	friend, people, interact, conversation, teacher, supportive, judge, family, advertise, impress, scene, parents, 'look smart', individual, peers, expectations, encouragement

to understand mathematics or a desire to avoid understanding, these responses provided evidence of cognitive processes related to mastery orientation.

Similarly, there was evidence that the performance-oriented items invoked constructs associated with performance orientation. When discussing doing better than other students in the class, Thelma stated, “It’s always nice to say I was in the group that got the A’s”. This social connotation associated with performance orientation and being part of a group appeared in many of the comments concerning performance; however, the students in this study may have social norms that are unique to this population. As Bill stated, “I don’t really focus on what the other students are doing ... but the grade will ultimately reflect who’s been doing better”. One of the items under review included the phrase *looking smart*. These same respondents took issue with this term. As Louise stated, “I don’t like to look smart ... [but to] be as smart as other people think I look, that would be nice”. For these students, fitting in was important, and looking smart and being better than others may have had negative connotations. Although, social comparisons were generally viewed negatively, these responses did lend evidence that the items referenced cognitive processes associated with performance orientation.

Items associated with expectancy generally invoked cognitive processes associated with self-efficacy and expectations of success. As Thelma stated, “Right now I’m doing very well in my math class ... doing all my homework, studying, putting in time and effort ... [and] if I keep that up I know that I will get a good grade”, and as Ted stated, “[If] I do put all my active effort into it, I definitely will [master the concepts]”. In general, responses to the expectancy items referenced cognitive processes associated with expectancy.

To verify content, we computed effect sizes defined as the proportion of concept phrases in our intra-respondent matrix that exemplify a code (Onwuegbuzie 2003). See Table 3 for the effect size per code for each item contained in the final MMAI. This revealed that the intrinsic motivation items tended to provoke cognitive processes associated with fun, efficacy, engagement, and affect. The mastery-oriented items evoked cognitive processes associated with relevance and mastery of content. The performance-oriented items evoked cognitive responses associated with performance comparisons and relationships. The expectancy items were associated with students’ perceived efficacy surrounding their classroom experiences.

Based on analyses of these interviews, we removed 15 items that did not reveal evidence of being representative of their intended construct. In this way, using an evidence-based approach, we narrowed in on four items each to represent intrinsic motivation, performance orientation, mastery orientation, self-efficacy, and expectancy, and three items each to represent task value and extrinsic motivation. Table 4 lists the reasons for the removal of the intrinsic motivation, performance orientation, mastery orientation, and expectancy items.

Table 3 Summary of intra-respondent matrix

	Fun	Master	Perform	Relevant	Efficacy	Engage	Affect	Related	Class	Wording
<i>Intrinsic motivation</i>										
im 1	0.13	0.07	0.00	0.22	0.00	0.24	0.07	0.00	0.02	0.09
im 2	0.24	0.02	0.00	0.00	0.18	0.20	0.31	0.00	0.07	0.09
im 3	0.31	0.14	0.03	0.00	0.21	0.17	0.14	0.00	0.03	0.07
im 4	0.17	0.05	0.05	0.07	0.07	0.01	0.02	0.01	0.01	0.14
<i>Mastery orientation</i>										
mo 1	0.10	0.19	0.06	0.21	0.10	0.00	0.00	0.04	0.19	0.04
mo 2	0.05	0.31	0.10	0.12	0.07	0.02	0.02	0.02	0.26	0.19
mo 3	0.06	0.26	0.03	0.26	0.34	0.09	0.03	0.09	0.00	0.17
mo 4	0.03	0.35	0.16	0.16	0.21	0.10	0.06	0.05	0.13	0.10
<i>Performance orientation</i>										
po 1	0.09	0.03	0.31	0.00	0.20	0.20	0.09	0.31	0.14	0.03
po 2	0.03	0.03	0.33	0.09	0.14	0.09	0.08	0.20	0.11	0.05
po 3	0.06	0.02	0.52	0.04	0.06	0.06	0.09	0.22	0.04	0.02
po 4	0.12	0.00	0.45	0.06	0.00	0.00	0.06	0.15	0.09	0.24
<i>Expectancy</i>										
ex 1	0.03	0.58	0.00	0.00	0.16	0.11	0.00	0.03	0.32	0.21
ex 2	0.00	0.10	0.45	0.05	0.15	0.45	0.00	0.05	0.25	0.00
ex 3	0.04	0.32	0.04	0.00	0.29	0.29	0.07	0.04	0.11	0.18
ex 4	0.00	0.06	0.52	0.03	0.06	0.00	0.00	0.18	0.15	0.03

Overall effect sizes per code per item. Unit of measure is a concept phrase

Key: intrinsic motivation (im), mastery orientation (mo), performance orientation (po), and expectancy (ex)

Phase 4: preliminary instrument

IRT for preliminary instrument

To begin an analysis using IRT, we administered this 26-item survey to 186 undergraduate college students taking their first math course at a public university in the southern part of the United States. Assuming all these items were measuring motivation, we compared a model with the item loadings constrained to be equal to a model where the item loadings were not constrained. The difference in the chi-square fit indices was significant $X^2(25)=111.765$, $p<0.0001$ implying the item loadings were significantly different. We analysed the item information curves for the unconstrained model, as well as the item characteristic curves. We found the information curves for the items representing self-efficacy, extrinsic motivation, and performance orientation were flat, providing little information for measuring the overall construct. The item characteristic curves for these poor performing items also revealed smooth low overlapping hills revealing no delineation between

Table 4 Evidence supporting removal of items from the four-factor Motivation for Mathematics Abbreviated Instrument (MMAI)

Item removed	Removal reason	Supporting data
<i>Intrinsic motivation</i>		
I do mathematics because I like the feeling of being totally immersed in the activity	<i>feeling immersed</i> was confusing and linked to mastery	T2-“if it’s something that they understand, they you will feel good, but if it’s something that they think they don’t understand and that they should be understanding then they just become totally... like don’t want to do it because I don’t know how to do it”.
I enjoy doing mathematics very much	<i>very much</i> was subjective	S1-“I enjoy doing math, [but] not very much”.
<i>Mastery orientation</i>		
I like mathematics that I’ve learned from even if I make a lot of mistakes	<i>a lot of mistakes</i> may have been counterproductive	T2-“from the student point of view making a lot of mistakes doesn’t make it fun”, “if I just keep going and I’m erring every time I’m just going to get frustrated”.
It’s important to me that I improve my math skills this year	<i>improvement</i> implied performance	S2-“my teachers didn’t care as much to recognize effort, they were just hoping for the students to pass”. “My teacher will recognize but for homework, quizzes, tests, etcetera”.
An important reason why I do mathematics is because I want to get better at it	<i>getting better</i> implied performance	S4-“I really really want to get a good grade”. “So that when it comes time for the test I don’t feel like I have to cram or sweat”.
<i>Performance orientation</i>		
In my math courses one of my goals has been to look smart in comparison to other students	<i>Look smart</i> had a negative social connotation	S1-“I don’t like to look smart”. S2-“I don’t really care what people think of me”. S4-“[I’m] not a show off”.
<i>Expectancy</i>		
I’m certain I can understand the most difficult material presented in a math class	Depends on how students defined difficult	T2-“They might be fractions is difficult and I’m like, ‘No, it’s not’”. S4-“I don’t know it is only difficult if you don’t understand it. If other people find it difficult then you think it is not difficult then you can’t really call it difficult”.

The sources of the data for the four students are labelled S1, S2, S3, and S4, and the two instructors are labelled T1 and T2

Fig. 1 Information curves both with and without poor performing items

Fig. 2 Initial eigenvalue plot with Parallel Analysis using preliminary data (n = 186)

answers. See Fig. 1 for the item information curves both with and without inclusion of the poor performing items.

EFA for preliminary instrument

We then conducted an EFA of the preliminary instrument. A plot of initial eigenvalues revealed a steepening of the slope between the fourth and fifth eigenvalue, and parallel analysis revealed the fourth eigenvalue for the actual data above and the fifth

Table 5 Geomin rotated pattern matrix for exploratory factor analysis of preliminary data

Item	1		2		3		4	
	est	SE	est	SE	est	SE	est	SE
im1	0.829	0.036	0.127	0.059	− 0.030	0.025	0.033	0.029
im2	0.743	0.045	0.114	0.052	0.002	0.020	− 0.080	0.074
im3	0.890	0.048	0.031	0.064	− 0.023	0.023	0.084	0.044
im4	0.939	0.036	− 0.013	0.058	0.053	0.048	0.010	0.041
mo1	0.131	0.058	0.807	0.064	0.012	0.056	− 0.073	0.071
mo2	− 0.002	0.042	0.902	0.048	− 0.033	0.062	− 0.027	0.056
mo3	0.053	0.075	0.788	0.061	− 0.028	0.067	0.050	0.066
mo4	0.045	0.075	0.753	0.072	− 0.028	0.066	0.035	0.062
em1	− 0.021	0.067	0.471	0.109	0.321	0.073	− 0.044	0.098
em2	− 0.146	0.105	0.213	0.112	0.320	0.099	− 0.094	0.091
em3	0.086	0.093	0.466	0.098	0.350	0.084	0.006	0.067
po1	0.115	0.124	− 0.123	0.092	0.749	0.075	0.015	0.060
po2	− 0.137	0.123	0.042	0.056	0.820	0.053	0.115	0.074
po3	− 0.083	0.114	0.146	0.079	0.818	0.058	0.056	0.074
po4	0.155	0.138	− 0.088	0.101	0.891	0.063	− 0.109	0.069
se1	0.147	0.078	− 0.085	0.066	0.031	0.051	0.773	0.060
se2	0.186	0.093	− 0.116	0.110	− 0.030	0.090	0.417	0.084
se3	0.413	0.109	− 0.006	0.128	0.132	0.116	− 0.047	0.100
se4	0.391	0.111	0.039	0.095	0.307	0.089	0.112	0.090
tv1	− 0.069	0.064	0.680	0.089	0.056	0.087	0.166	0.093
tv2	0.246	0.085	0.472	0.078	− 0.020	0.078	0.177	0.084
tv3	0.030	0.075	0.734	0.073	0.023	0.069	0.046	0.073
ex1	− 0.036	0.054	0.086	0.079	− 0.021	0.050	0.798	0.062
ex2	0.065	0.058	− 0.039	0.040	0.009	0.057	0.883	0.042
ex3	− 0.001	0.045	0.130	0.086	0.104	0.053	0.774	0.056
ex4	− 0.023	0.039	0.058	0.045	− 0.027	0.044	0.914	0.040

Key: intrinsic motivation (im), mastery orientation (mo), performance orientation (po), extrinsic motivation (em), self-efficacy (se), task value (tv), and expectancy (ex)

$n=186$

Table 6 Items contained in the four-factor MMAI

Item #	Item content
im 1	I would describe mathematics as very interesting
im 2	I do mathematics for the excitement I feel when I am really involved in the activity
im 3	I think mathematics is enjoyable
im 4	Mathematics is fun to do
mo 1	It's important to me that I learn a lot of new math concepts this year
mo 2	One of my goals is to master a lot of new mathematics this year
mo 3	It's important to me that I thoroughly understand mathematics
mo 4	One of my goals in my math courses has been to learn as much as I can
po 1	I have liked showing math teachers that I'm smarter than the other students
po 2	I have felt successful in my math courses when I did better than the other students
po 3	In math courses, I have wanted to do better than other students
po 4	Doing better than other students in my math courses has been important to me
ex 1	I'm confident I can understand the basic concepts taught in this math class
ex 2	I believe I will receive excellent grades in this math class
ex 3	I'm certain I can master the skills being taught in this math class
ex 4	I expect to do well in this math class

*Key:*intrinsic motivation (im), mastery orientation (mo), performance orientation (po), and expectancy (ex)

eigenvalue below the eigenvalues from the simulated random data. Thus, the data are best fit with a four-factor model. See Fig. 2 for eigenvalue plot.

Analysis of the rotated pattern matrix revealed the intrinsic motivation, mastery orientation, performance orientation, and expectancy items separating cleanly into the four factors where within factor loadings were all greater than 0.74 and all loadings across factors being less than 0.16. The extrinsic motivation and self-efficacy items loaded across factors, and therefore we removed them from the instrument. We also removed the task-value items as they weakly loaded on one factor. See Table 5 for the pattern matrix. The items representing intrinsic motivation, mastery orientation, and expectancy performed best under the IRT analysis, and these items as well as the items representing performance orientation revealed a strong four-factor structure. As a result of theoretical concerns and these statistical results, we focussed the MMAI on a four-factor model with the four factors being intrinsic motivation, mastery orientation, performance orientation, and expectancy. Table 6 lists the items contained in the final 16-item four-factor MMAI.

Phase 5: internal structure of the MMAI

We administered the 16-item MMAI to 386 university participants in general education mathematics courses offered from the fall of 2015 through the summer of 2017 at a public university in the southern United States. Of these participants, 211

identified as females, 167 as males, and eight students identified as other or choose not to answer.

CFA for final instrument

We conducted a CFA of the four-factor 16-item MMAI, and although the chi-square value of $X^2(98, n=386)=212.02$, $p<0.0001$ indicated misfit, other fit indices less sensitive to sample size revealed a good fit as per levels reported by Browne and Cudeck (1993), MacCallum et al. (1996), and Hu and Bentler (1999). The four-factor model had an RMSEA=0.055. The SRMR was 0.044, and the CFI was 0.963. All indicate a good fit.

All other sources of misfit with modification indices greater than 10 concerned correlated errors of items within factors. Some expectancy items had correlated errors. Ex3 with Ex1 had a modification index of $X^2(1)=12.70$, $p<0.0001$. Ex4 with Ex2 had a modification index of $X^2(1)=16.12$, $p<0.0001$. Some mastery items had correlated errors for Mo2 with Mo1 and Mo4 with Mo1. Their modification indices were $X^2(1)=14.39$, $p<0.0001$ and $X^2(1)=10.43$, $p<0.0001$, respectively. Two performance orientation items—Po1 and Po3—also had correlated error with a modification index of $X^2(1)=12.62$, $p<0.0001$. Modification indices showed some correlated error between items within the same latent factors, however, in view of the good fit of the original model, none of these sources of misfit warranted changing the model. See Table 7 for fit indices and modification indices.

Internal reliability and validity

To examine the internal reliability of the measurement model, we found the composite reliability for all four latent factors to be excellent with a lowest of 0.86 and the highest at 0.93. These are well above Fornell and Larcker's (1981) minimum threshold of 0.70. To examine convergent validity in the measurement model, we compute AVE for each latent construct. These values were also excellent as the lowest of 0.60 was well above Fornell and Larcker's (1981) minimum threshold of 0.50. To examine discriminate validity in the measurement model, we compared the

Table 7 Fit indices and modification indices for CFA

Fit indices	df	χ^2	<i>p</i> value	CFI	RMSEA	SRMR
Four-factor model	98	212.02	<0.0001	0.963	0.055	0.044
Modification indices	df	χ^2	<i>p</i> value			
Ex2 with Ex4	1	16.12	<0.0001			
Mo1 with Mo2	1	14.39	<0.001			
Ex1 with Ex3	1	12.70	<0.001			
Po1 with Po3	1	12.62	<0.001			
Mo1 with Mo4	1	10.43	<0.01			

Table 8 Summary table for reliability and validity estimates for the measurement model

	$\rho_{AVE}(\xi_j)$	$\rho_{CR}(\xi_j)$	Λ_i	ρ_{ξ_i}	mo	po	ex
im	0.766	0.929			$\gamma_{im\ mo} = 0.489$	$\gamma_{im\ po} = 0.255$	$\gamma_{im\ ex} = 0.419$
			λ_{im1}	0.845	0.71		
			λ_{im2}	0.798	0.64		
			λ_{im3}	0.923	0.85		
			λ_{im4}	0.931	0.87		
mo	0.642	0.877				$\gamma_{mo\ po} = 0.218$	$\gamma_{mo\ ex} = 0.351$
			λ_{mo1}	0.805	0.65		
			λ_{mo2}	0.830	0.69		
			λ_{mo3}	0.780	0.61		
			λ_{mo4}	0.779	0.61		
po	0.603	0.858					$\gamma_{po\ ex} = 0.314$
			λ_{po1}	0.660	0.44		
			λ_{po2}	0.798	0.64		
			λ_{po3}	0.827	0.68		
			λ_{po4}	0.818	0.67		
ex	0.673	0.891					
			λ_{ex1}	0.759	0.58		
			λ_{ex2}	0.845	0.71		
			λ_{ex3}	0.814	0.66		
			λ_{ex4}	0.861	0.74

*All relationships are significant with $p < .001$ **Fig. 3** Results from confirmatory factor analysis. *All relationships are significant with $p < 0.001$

squared correlations between factor loadings to AVE with the requirement that $\rho_{\text{AVE}(\xi_j)} > \gamma^2$, where γ is the correlation of the factor ξ_j with any other factor. All four factors met this requirement. All individual item reliabilities were above Ab Hamid et al.'s (2017) minimum threshold of 0.4, however, the first performance orientation item was concerning with $\rho_{\text{po1}} = 0.44$. See Table 8 for the complete summary.

Significant relationships

All item loadings on the latent factors were significantly different from zero with $p < 0.001$. The range of the standardized loadings for the intrinsic factor was 0.80 to 0.93, the range for the mastery orientation factor was 0.78 to 0.83, the range for the performance orientation factor was 0.66 to 0.83, and the range for the expectancy factor was 0.76 to 0.86. All correlations between latent factors were also significantly different than zero with $p < 0.001$. Correlations between the factors ranged from 0.22 (performance and mastery orientation) to 0.49 (intrinsic and mastery orientation). These low-factor inter-correlations provided evidence of discriminant validity between factors. See Fig. 3 for the CFA model.

Measurement invariance

We conducted invariance testing grouping students as male and female. Starting with a configural model, we successively added constraints and compared the successive fit indices. For the metric model, we constrained the factor loading to be the same across groups. For the scalar model we also constrained the intercepts to be the same, and finally, for the strict model, both item level and latent factor level errors were constrained to be the same. Using the Satorra and Bentler (2010) scaled chi-square differences, we only found significant change in fit between the metric and scalar models ($X^2(12) = 26.35$, $p = 0.01$). RMSEA stayed remarkably stable fluctuating between 0.058 and 0.059 throughout this entire process, however, CFI and SRMR showed some degradation from the added constraints with CFI moving from 0.958 to 0.951 and SRMR moving from 0.053 to 0.071. See Table 9 for a summary of the invariance testing data.

Table 9 Summary table for measurement invariance

	df	χ^2	* $\Delta\chi^2$	Δdf	p -value	CFI	RMSEA	SRMR
Configural	196	324.7				0.958	0.059	0.053
Metric	208	341.5	16.40	12	0.173	0.957	0.058	0.058
Scalar	220	367.1	26.35	12	0.010	0.952	0.059	0.06
Strict	236	388.3	23.02	16	0.113	0.951	0.058	0.071

Summary table of likelihood ratio test for differential item functioning across gender using Mplus version 7.4 and MLR estimator.

*Satorra and Bentler (2010) scaled chi-square differences were used

Discussion

Our goal was to develop a brief motivational measure grounded in theory that could be used for applied research in mathematics education. We began with items from well-established measures that have strong theoretical and psychometric support, and then used a combination of qualitative and quantitative methods to identify a manageable set of items to measure four motivational constructs: intrinsic motivation, mastery orientation, performance orientation, and expectancy. Input from education researchers interested in motivation along with feedback from students who shared their interpretations of the meaning of each item and their thought processes used in answering the items, served to verify the content of the items for a college-student population. Despite the fact that the motivation items came from previously vetted questionnaires, we used this process to verify the item's content was not significantly changed in our process, and was relevant for these students. We build on the previous evidence of content validity of the MMAI items and provide additional evidence that the items in the MMAI represent their intended constructs.

Feedback from educational researchers and students was used to narrow the number of items to a manageable set. By using multiple sources and types of data (quantitative and qualitative), this triangulation added to the richness of the development and validation process, and by quantifying the qualitative data we provided some control for personal bias.

The purpose of the MMAI is to measure student motivation for mathematics across four constructs. The main focus of the first two stages of this study was to select the items that best represented the constructs for this purpose. All the items had been used extensively in social science and educational research to represent their intended construct so they have some inherent validity of content. The survey of educational researchers interested in motivation helped narrow in on the most representative items, and interviews with students helped verify that the items were still aligned to their intended construct. The results from these first two stages suggest they were. These results should seem unremarkable as these items have been used hundreds of times in research to represent these exact constructs.

Next, we completed a trial run of a preliminary instrument with items representing seven different constructs and conducted quantitative analyses of the responses. We used IRT to examine how the items were functioning individually and collectively in a single factor model. By examining differences in fit between a constrained (equal item slopes) and an unconstrained model (varying item slopes), we found an unconstrained model to be appropriate. This implied the items were performing differently. A pattern emerged when examining the item response and information curves. We found all items representing extrinsic motivation, performance orientation, and self-efficacy were performing poorly and providing minimal information about the overall construct.

We then utilized EFA to determine the factor structure within this preliminary data. We employed Kaiser's criterion (eigenvalues greater than or equal to 1.0); parallel analysis, which involves retaining a factor if the i th eigenvalue from the actual data is larger than the i th eigenvalue from simulated random data; and visual

analysis of the scree plot of the eigenvalues (i.e. a noticeable steepening of the slope between two eigenvalues indicating an end to the number of necessary factors), to reveal a strong four-factor structure. Assuming that these four factors were correlated, we used Geomin rotation and examined the pattern matrix to find a strong separation along theoretical lines for the intrinsic motivation, mastery orientation, performance orientation, and expectancy items into four distinct factors. Extrinsic motivation and self-efficacy items loaded weakly across factors. This may be a result of the multidimensional nature of these constructs. Extrinsic motivation is framed as a continuum with four designations based on regulatory style. One's regulatory style can be external, introjected, identified, or integrated. The self-efficacy items were rooted in the sources of self-efficacy: mastery experiences, social persuasion, vicarious experience, and emotional states. The faceted nature of these constructs may have caused them to perform poorly on an abbreviated instrument, and so we removed them. Finally, the task-value items did not load cleanly along theoretical lines tending to moderately load with the mastery orientation items. There also was a realization that the task-value items did not adequately represent the multi-faceted concept of value. Because of these reasons, we removed the task-value items.

Notice, in the IRT analysis of item information curves, we found the performance orientation items did not add information to understand the overall construct of motivation. In the EFA, these same items separated strongly into their own factor, and in the CFA, the performance orientation factor had the lowest correlations to other factors. It may be that performance orientation as a social construct behaves differently than the other three factors. Because of the strong discrimination in EFA and a desire to more fully represent achievement goal theory, and a realization for the importance of performance orientation in high-stakes testing environments, we kept the performance orientation items in the final instrument.

For the final phase, we demonstrated internal structural validity through a CFA. The 16-item four-factor measurement model had good fit to data from 386 post-secondary students in undergraduate mathematics courses. To determine how well a predetermined latent factor structure fit the data, we define the model, ran it, and examined various fit indices. Because of its bias against large sample sizes, the chi-square fit index will almost always show significant misfit, and therefore, it is generally not useful. Instead, for convergent validity we relied on SRMR as it is most sensitive to factor covariance misspecification, and we relied on RMSEA and CFI for discriminant validity as they are sensitive to factor loading misspecification. Discriminant validity is how distinct the factors are as measured by separate sets of items, and convergent validity is how cohesive the sets of items are in measuring their latent factor (Sun 2005). This model showed the items within each factor were strongly related providing evidence of convergence within factors, while the motivational factors were only moderately correlated with each other providing evidence of discriminant validity.

We examined the instrument using IRT and found the intrinsic motivation, mastery orientation, and expectancy items to provide almost all information about the overall construct. This relationship between these latent constructs was also revealed when we examined the instrument using CFA. Here, intrinsic motivation was moderately correlated with mastery orientations ($r = 0.49$) and expectancy ($r = 0.42$).

This relationship was expected as mastery orientations have been shown to have positive correlations with intrinsic motivation (Anderman et al. 1999), and expectancy and intrinsic motivation should be correlated as expectancy is a behavioural approach and intrinsic motivation is a cognitive approach to defining motivation. Interestingly, performance orientation had its highest correlation with expectancy ($r = 0.31$), and its lowest correlation with mastery orientations ($r = 0.22$). This may be reasonable as performance orientations are comparisons of behaviour, and performance orientations are also designed as a counterpoint in achievement goal theory to mastery orientations.

Higher scores on the MMAI for intrinsic motivation, mastery orientation, and expectancy are indicative of higher levels of motivation for mathematics, however, this is not necessarily true of higher scores on performance orientation. In high-stakes testing environments, performance orientations may be useful for understanding classroom interactions, but a student who scores high on the other three factors for motivation may or may not be performance oriented. Achievement goal theory frames academic motivations as performance or mastery oriented. Notice, mastery concerns understanding math, whereas performance concerns using math to compare oneself to others. Intrinsic motivation, mastery orientations, and expectancy have a direct relationship to the domain, whereas performance orientations are about behaviour within the domain. This difference in focus needs to be considered when analysing performance orientations as part of the four-factor MMAI model. The results of the CFA along with the excellent reliability coefficients provide strong support for the psychometric quality of the MMAI.

We also conducted measurement invariance testing across the two groups of students that identified as male or female. This does not imply a binary mindset towards human sexuality, instead it is meant to encourage research into the disproportionate representation of females and males in undergraduate mathematics and education programs. An instrument that is measurement invariant is structurally unbiased between groups. The same fit indices used in CFA are relevant here, and although chi-square does have a bias against large sample size, chi-square differences between competing models are often reported. Despite this, the other fit indices demonstrated good fit throughout the measurement invariance testing, and these results provide preliminary support for gender invariance of the MMAI.

The evidence presented makes a strong argument that the MMAI does meet its intended purpose. The items represent their intended constructs, and the measurement model underlying the instrument provides a good fit to the data. We confirmed the validity of the content through expert review, and student and teacher commentary. We used IRT and EFA to narrow in on four constructs representing theories of motivation in education, and we determined that there were no theoretical or statistically meaningful competing structures. Then we used CFA to demonstrate the model was a good fit to the data, and invariance testing to demonstrate the instrument may be used to measure differences between genders. Taken together, these results support the validity of the MMAI, which is important because researchers need tools that can be implemented quickly in educational settings to measure motivation for mathematics, and they need to have trust that the instrument is performing as expected.

This study contributes to the field of educational research by completing much of the work necessary to provide a brief instrument, aligned with motivation theory, that has strong psychometric properties. These attributes should promote research that is based in theory. The brevity of the instrument should facilitate researchers' use of the measure, and its alignment to theory promotes advancement of motivational theory. By allowing for more measurements in more situations, researchers may better understand the drivers of motivation. In this vein, additional research is needed to evaluate the applicability of the MMAI to a wider audience. Work with students in secondary as well as other post-secondary mathematics courses would be beneficial. We also encourage research evaluating the psychometric properties of the MMAI in longitudinal research to evaluate changes in student motivation over time. One reason for developing this brief motivational instrument was to provide researchers with a practical tool that could be used periodically within the constraints of classroom schedules.

Additional research comparing other related constructs (e.g. self-efficacy, task value, and anxiety) as part of the nomological network of motivation is also of interest. The MMAI focuses on achievement goals framed by the predominant cognitive and behavioural theories of motivation; however, we understand this may not meet the demands of all who are interested in educational research.

In conclusion, the MMAI, as a measure of four motivational constructs (intrinsic motivation, mastery orientation, performance orientation, and expectancy), has strong evidence of content and internal structure validity. The MMAI has the added benefit of being theoretically grounded and brief, which makes it a practical and useful tool for classroom-based research into motivation.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s43545-021-00276-6>.

Author contributions KB conceived the study, collected and analysed the data, and wrote the initial draft of the paper. RFD verified the analytical methods and provided critical feedback in the design of the study and writing the manuscript.

Data availability Data and code available at osf.io under Content and Internal Structural Validity of the Motivation for Mathematics Abbreviated Instrument.

Declarations

Conflict of interests The authors declare that they have no conflict of interests.

Ethical Approval On 5/22/2015, the Institutional Review Board (IRB) of the University of South Florida determined that this research meets criteria for exemption from the federal regulations as outlined by 45CFR46.101(b).

Informed consent An informed consent form was explained to the participants just before they took the survey. This explained that participation was voluntary, anonymous, and they had to be over 18 years old. No individual identifying records were kept.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative

Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ab Hamid MR, Sami W, Sidek MM (2017) Discriminant validity assessment: use of Fornell & Larcker criterion versus HTMT criterion. *J Phys* 890(1):012163
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (2014) Standards for educational and psychological testing. American Educational Research Association, Washington
- Anderman EM, Maehr ML, Midgley C (1999) Declining motivation after the transition to middle school: schools can make a difference. *J Res Dev Educ* 32:131–147
- Bagozzi RP, Yi Y (1988) On the evaluation of structural equation models. *J Acad Mark Sci* 16(1):74–94
- Bandura A (1997) Self-efficacy: the exercise of self-control. WH Freeman & Co, New York
- Browne MW, Cudeck R (1993) Alternative ways of assessing model fit. *Sage Sociol Methods Res* 21(2):230–258. <https://doi.org/10.1177/00491241920021002005>
- Coelho GLH, Monteiro RP, Hanel PHP, Vilar R, Gouveia VV, Maio GR (2018) Psychometric parameters of an abbreviated vengeance scale across two countries. *Personal Individ Differ* 120:185–192. <https://doi.org/10.1016/j.paid.2017.08.042>
- Costello AB, Osborne J (2005) Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Pract Assess Res Eval* 10(1):185–192. <https://doi.org/10.1016/j.paid.2017.08.042>
- Cronbach LJ, Meehl PE (1955) Construct validity in psychological tests. *Psychol Bull* 52(4):281–302. <https://doi.org/10.1037/h0040957>
- Deci E, Ryan R (1985) Intrinsic motivation and self-determination in human behavior. Plenum Publishing Corp, New York
- Eccles JS (1987) Gender roles and women's achievement-related decisions. *Psychol Women Q* 11(2):135–172. <https://doi.org/10.1111/j.1471-6402.1987.tb00781.x>
- Eccles JS, Wigfield A (1995) In the mind of the actor: the structure of adolescents' achievement task-values and expectancy-related beliefs. *Pers Soc Psychol Bull* 21(3):215–225. <https://doi.org/10.1177/0146167295213003>
- Elliot AJ, Harackiewicz JM (1996) Approach and avoidance achievement goals and intrinsic motivation: a meditational analysis. *J Pers Soc Psychol* 70(3):461–475. <https://doi.org/10.1037/0022-3514.70.3.461>
- Fornell C, Larcker DF (1981) Evaluating structural equation models with unobservable variables and measurement error. *J Mark Res* 18(1):39–50
- Gogol K, Brunner M, Goetz T, Martin R, Ugen S, Keller U, Fischbach A, Preckel F (2014) "My Questionnaire is Too Long!" The assessments of motivational-affective constructs with three-item and single-item measures. *Contemp Educ Psychol* 39(3):188–205. <https://doi.org/10.1016/j.cedpsych.2014.04.002>
- Hancock D (2007) Exploring the effects of group testing on graduate students' motivation and achievement. *Assess Eval High Educ* 32(2):215–227. <https://doi.org/10.1080/02602930601051176>
- Hu L, Bentler PM (1999) Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model* 6(1):1–55. <https://doi.org/10.1080/10705519909540118>
- Kosovich J, Hulleman C, Barron K (2017) Measuring motivation in educational settings: a case for pragmatic measurement. In: Renninger KA, Hidi SE (eds) *The Cambridge handbook on motivation and learning*. 1st edn. Routledge, London, pp 39–60

- Li L (2017) The role of anonymity in peer assessment. *Assess Eval High Educ* 42(4):645–656. <https://doi.org/10.1080/02602938.2016.1174766>
- MacCallum RC, Browne MW, Sugawara HM (1996) Power analysis and determination of sample size for covariance structure modeling. *Psychol Methods* 1(2):130–149. <https://doi.org/10.1037/1082-989X.1.2.130>
- McAuley E, Duncan T, Tammen VV (1987) Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis. *Res Q Exerc Sport* 60(1):48–58. <https://doi.org/10.1080/02701367.1989.10607413>
- Meredith W (1993) Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 58(4):525–543. <https://doi.org/10.1007/BF02294825>
- Midgley C, Maehr ML, Hruda LZ, Anderman E, Anderman L, Freeman KE, Urdan T (2000) Manual for the patterns of adaptive learning scales (PALS). University of Michigan, Michigan
- Miller RB, DeBacker TK, Greene BA (1999) Perceived instrumentality and academics: the link to task valuing. *J Instr Psychol* 26(4):250–350
- Monteiro V, Mata L, Peixoto F (2015) Intrinsic Motivation Inventory: psychometric properties in the context of first language and mathematics learning. *Psicologia* 28(3):434–443. <https://doi.org/10.1590/1678-7153.201528302>
- Muthén LK, Muthén BO (1998–2017) Mplus user's guide, Eighth Edition. Muthén & Muthén, Los Angeles
- Onwuegbuzie AJ (2003) Effect sizes in qualitative research: a prolegomenon. *Qual Quant* 37:393–409. <https://doi.org/10.1023/A:1027379223537>
- Onwuegbuzie AJ, Teddlie C (2003) A framework for analyzing data in mixed methods research. In: Tashakkori A, Teddlie C (eds) *Handbook of mixed methods in social and behavioral research*, 2nd edn. Sage, Thousand Oaks, pp 397–430
- Pelletier LG, Fortier MS, Vallerand RJ, Tuson KM, Briere NM, Blais MR (1995) Toward a new measure of intrinsic motivation, extrinsic motivation, and amotivation in sports: the Sport Motivation Scale (SMS). *J Sport Exerc Psychol* 17(1):35–53. <https://doi.org/10.1123/jsep.17.1.35>
- Pelrman D (2010) Change in affect and needs satisfaction for amotivated students within the sport education model. *J Teach Phys Educ* 29(4):433–445
- Pintrich PR, Smith D, Garcia T, McKeachie WJ (1992) A manual for the use of the motivated strategies for learning questionnaire (MSLQ). Office of Educational Research and Improvement, Washington
- Putnick DL, Bornstein MH (2016) Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Dev Rev* 41:71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Rammstedt B, Beierlein C (2014) Can't we make it any shorter? *J Individ Differ* 35:212
- Ratelle CF, Guay F, Vallerand RJ, Larose S, Senécal C (2007) Autonomous, controlled, and amotivated types of academic motivation: a person-oriented analysis. *J Educ Psychol* 99(4):734
- Ryan R, Deci E (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am Psychol* 55(1):68–78. <https://doi.org/10.1037/0003-066x.55.1.68>
- Ryan RM, Kuhl J, Deci EL (1997) Nature and autonomy: an organizational view of social and neurobiological aspects of self-regulation in behavior and development. *Dev Psychopathol* 9(4):701–728. <https://doi.org/10.1017/s0954579497001405>
- Samejima F (1996) The graded response model. In: van der Linden WJ, Hambleton RK (eds) *Handbook of modern item response theory*. Springer, Berlin
- Satorra A, Bentler PM (2010) Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika* 75:243–248. <https://doi.org/10.1007/s11336-009-9135-y>
- Schmitt N, Kuljanin G (2008) Measurement invariance: review of practice and implications. *Hum Resour Manag Rev* 18(4):210–222. <https://doi.org/10.1016/j.hrmr.2008.03.003>
- Simper N, Kaupp J, Frank B, Scott J (2016) Development of the transferable learning orientation tool: providing metacognitive opportunities and meaningful feedback for students and instructors. *Assess Eval High Educ* 41:1159–1175. <https://doi.org/10.1080/02602938.20105.1070117>
- Strauss AL, Corbin JM (1990) *Basics of qualitative research*, vol 15. Sage, Thousand Oaks
- Sun J (2005) Assessing goodness of fit in confirmatory factor analysis. *Meas Eval Couns Dev* 37(4):240–256
- Tourangeau R (1984) Cognitive science and survey methods: a cognitive perspective. In: Jabine H, Straf M, Tanur J, Tourangeau R (eds) *Cognitive aspects of survey methodology: building a bridge between disciplines*. National Academy Press, Washington, pp 73–100

- Usher EL, Pajares F (2009) Sources of self-efficacy in mathematics: a validation study. *Contemp Educ Psychol* 34(1):89–101. <https://doi.org/10.1016/j.cedpsych.2008.09.002>
- Vallerand RJ, Pelletier LG, Blais MR, Briere NM, Senecal C, Vallieres EF (1992) The academic motivation scale: a measure of intrinsic, extrinsic, and amotivation in education. *Educ Psychol Meas* 52(4):1003–1017
- Vandenberg RJ, Lance CE (2000) A review and synthesis of the MI literature: suggestions, practices, and recommendations for organizational research. *Organ Res Methods* 3:4–69. <https://doi.org/10.1177/1094428100310>
- Wibrowski CR, Matthews WK, Kitsantas A (2017) The role of a skills learning support program on first-generation college students' self-regulation, motivation, and academic achievement: a longitudinal study. *J Coll Stud Retent* 19(3):317–332. <https://doi.org/10.1177/1521025116629152>
- Willis GB (2004) *Cognitive interviewing: a tool for improving questionnaire design*. Sage, Thousand Oaks
- Yin H, Wang W (2016) Undergraduate students' motivation and engagement in China: an exploratory study. *Assess Eval High Educ* 41(4):601–621. <https://doi.org/10.1080/02602938.2015.1037240>