

Portland State University

PDXScholar

---

Chemistry Faculty Publications and  
Presentations

Chemistry

---

7-1-2020

# Investigation of Evidence for the Internal Structure of a Modified Science Motivation Questionnaire II (msmq II): a Failed Attempt to Improve Instrument Functioning Across Course, Subject, and Wording Variants

Regis Komperda  
*San Diego State University*

Kathryn Hosbein  
*Portland State University*

Michael M. Phillips  
*University of Northern Colorado*

Jack Barbera  
[jackbarbera@pdx.edu](mailto:jackbarbera@pdx.edu) [pdxscholar.library.pdx.edu/chem\\_fac](https://pdxscholar.library.pdx.edu/chem_fac)

 Part of the [Chemistry Commons](#), and the [Science and Mathematics Education Commons](#)

Let us know how access to this document benefits you.

---

## Citation Details

Komperda, R., Hosbein, K. N., Phillips, M. M., & Barbera, J. (2020). Investigation of evidence for the internal structure of a modified science motivation questionnaire II (mSMQ II): a failed attempt to improve instrument functioning across course, subject, and wording variants. *Chemistry Education Research and Practice*, 21(3), 893–907. <https://doi.org/10.1039/d0rp00029a>

This Post-Print is brought to you for free and open access. It has been accepted for inclusion in Chemistry Faculty Publications and Presentations by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

## **Investigation of evidence for the internal structure of a modified Science Motivation Questionnaire II (mSMQ II): A failed attempt to improve instrument functioning across course, subject, and wording variants**

<sup>a</sup>Regis Komperda, <sup>b</sup>Kathryn N. Hosbein, <sup>c</sup>Michael M. Phillips, and <sup>b\*</sup>Jack Barbera

<sup>a</sup>Department of Chemistry & Biochemistry; Center for Research in Mathematics and Science Education, San Diego State University

<sup>b</sup>Department of Chemistry, Portland State University

<sup>c</sup>School of Psychological Sciences, University of Northern Colorado

\*jbarbera@pdx.edu

### **Abstract**

The Science Motivation Questionnaire II (SMQ II) was developed to measure aspects of student motivation in college-level science courses. Items on the SMQ II are structured such that the word ‘science’ can be replaced with any discipline title (e.g., chemistry) to produce a discipline-specific measure of student motivation. Since its original development as the Science Motivation Questionnaire and subsequent refinement, the SMQ II and its discipline-specific variants have been used in a number of science education studies. However, many studies have failed to produce acceptable validity evidence for their data based on the proposed internal structure of the instrument. This study investigated if modifications could be made to the SMQ II such that it produces consistent structural evidence across its use in various forms. A modified SMQ II (mSMQ II) was tested with wording variants (‘science’ and ‘biology’ or ‘chemistry’) in general biology and in preparatory and general chemistry courses at several institutions. Exploratory and confirmatory factor analysis were used to cull problematic items and evaluate the structure of the data based on the relations posited by the SMQ II developers. While extensive revisions resulted in acceptable data model fit for the five-factor structural models in most course and wording conditions, significant issues arose for the single-factor scales. Therefore, potential users are cautioned about the utility of the SMQ II or its variants to support the evaluation of classroom

practices. A reflective review of the theoretical underpinnings of the SMQ II scales call into question the original framing of the scales and suggests potential alternatives for consideration.

**Reference Information:**

Regis Komperda, Kathryn N. Hosbein, Michael M. Phillips, and Jack Barbera, "Investigation of evidence for the internal structure of a modified Science Motivation Questionnaire II (mSMQ II): A failed attempt to improve instrument functioning across course, subject, and wording variants," *Chemistry Education Research and Practice*, 2020, 21, 893-907, DOI:10.1039/D0RP00029A.

## Introduction

The discipline-based education research community continues to recognize the importance of including the affective domain when studying student outcomes in science courses (National Research Council, 2012; Fortus, 2014). Motivation is one aspect of the affective domain frequently investigated in the field of chemistry education research (Black and Deci, 2000; Zusho *et al.*, 2003; Chan and Bauer, 2014; González and Paoloni, 2015; Salta and Koulougliotis, 2015; Ferrell *et al.*, 2016; Ardura and Pérez-Bitrián, 2018; Austin *et al.*, 2018; Liu *et al.*, 2018) as well as in other science fields (Simpkins *et al.*, 2006; Glynn *et al.*, 2009; Olimpo *et al.*, 2016; Schumm and Bogner, 2016; González *et al.*, 2017; Young *et al.*, 2018; Zeyer, 2018). One commonality among all these studies is their use of self-report survey instruments for measuring student motivation.

When developing instruments to measure unobservable (i.e., latent) traits such as motivation, it is necessary to align the items on the instrument with a theoretical framework for the latent variable (American Educational Research Association *et al.*, 2014). In the case of motivation, the literature contains multiple theoretical frameworks including social-cognitive theory (Bandura, 1993), self-determination theory (Ryan and Deci, 2000), and expectancy-value theory (Wigfield and Eccles, 2000), among others. One instrument combining multiple motivation theories is the Science Motivation Questionnaire (SMQ; Glynn and Koballa, 2006; Glynn *et al.*, 2009), which was later revised by the developers into the Science Motivation Questionnaire II (SMQ II; Glynn *et al.*, 2011).

### Theoretical framework of the SMQ II

Glynn and colleagues' work (e.g., 2006; 2007, 2009, 2011) on the Science Motivation Questionnaire (SMQ and SMQ II) used self-regulation as the overarching framework of their five-factor motivation instrument which includes the individual motivation scales of intrinsic, extrinsic [grade and career], self-determination, self-efficacy, relevance, and anxiety. Yet, many of the scales do not address the unique aspects of how students self-regulate their thoughts, actions, environment, and motivation to achieve their academic goals (Zimmerman, 2000). As Eccles and Wigfield (2002) note upon review, self-regulated learning tends to include three important aspects: self-observation while engaged with an academic task, self-judgment regarding one's performance, and self-evaluation or reactions to one's performance after a task has been completed. Given the absence of items aligned with these aspects, a reflective evaluation is needed regarding the conceptual/theoretical underpinnings for the SMQ II in addition to the psychometric characteristics of the scales.

Within the SMQ II (Glynn *et al.*, 2011), the only items that specifically address aspects of self-regulation are on the self-determination scale (factor 3; p. 1167), as these items focus on study preparation and effort exertion for studying science (e.g., "I put enough effort into learning science" or "I prepare well for science tests and labs"). Additionally, the scale itself might not align with a framework of self-determination, particularly if, as described by the authors (p. 1161) this definition arises from the self-determination theory (SDT; Deci and Ryan, 2000), which centers more on the three psychological needs of autonomy, competence, and relatedness. When these three needs are met (to varying degrees), an individual's actions are more self-determined, which can influence regulatory styles (as described in SDT) across the extrinsic-intrinsic continuum. Self-determined actions are growth-oriented and are not overly impacted by external influences, which is how self-determined actions are related to the distinction between

intrinsic and extrinsic motivation. This continuum contrasts with the dichotomy implied by the separation of the intrinsic motivation scale from the extrinsic scale in the SMQ and later the extrinsic focused scales of grade and career motivation in the SMQ II. An additional concern regarding the theoretical framework of the SMQ II scales is the inclusion of self-determination as a distinct construct. Though Ryan and Deci describe their theory of motivation as self-determination theory (SDT), self-determination tends to describe motivated action when one's psychological needs are being met (Ryan and Deci, 2000) and can range on a continuum from extrinsic to intrinsic motivation rather than being a distinct construct.

The primary support for Glynn and colleagues' (2011) proposed theoretical framework for the SMQ II comes from analyses of the internal structure of the instrument using both exploratory and confirmatory factor analysis. Factor analysis statistical techniques allow researchers to determine if instrument data are aligned with a hypothesized internal structure, a form of model testing critical to the practice of science (Grosslight *et al.*, 1991). In the case of the SMQ II, this takes the form of a model containing five distinct yet related aspects of motivation in a five-factor model (intrinsic, extrinsic [grade and career], self-determination, self-efficacy). Results from exploratory factor analysis in prior work (Glynn *et al.*, 2009) showed that extrinsic motivation consisted of two separate but related components: grade and career motivation. When extrinsic motivation was split into these two components, the five-factor motivation model was shown to provide adequate fit to samples of students within major and non-major biology courses (Glynn *et al.*, 2011). While these two types of extrinsic motivation were supported through factor analysis, these results alone do not provide strong theoretical support for the new constructs. If subsequent data are found to poorly fit this model, or if the aspects of motivation measured by the SMQ II are not found to be distinct factors within

additional samples, this provides an opportunity to examine potential issues with the underlying model of motivation or the items developed to measure it and further refine the items and/or model. This model testing should occur at each use of the instrument to support the validity of the data collected and ensure the results can be interpreted in a meaningful way (American Educational Research Association *et al.*, 2014).

As assessment instruments are commonly used within the chemistry education community to provide insight to the impacts of classroom practice, it is imperative that the data produced by an assessment instrument shows evidence of validity and reliability (Arjoon *et al.*, 2013). Furthermore, if assessment items used to measure a relevant trait, such as self-efficacy, are not shown to align with a theory of self-efficacy, the interpretation of the results may not be reflective of a learning environment's support of, or impact on, the trait. Therefore, interpreting data from assessment instruments that do not show evidence of validity and reliability can lead to misinformed judgements about classroom practice. As the SMQ II is purported as an assessment tool that can be used across a range of courses and disciplines (Glynn *et al.*, 2011), data from its different wording variants and applications must be equally supported by evidence.

### **Prior studies of SMQ II internal structure**

The SMQ II is a revision of the SMQ by the original developers based on both student interviews and factor analysis of 1,450 student responses to original and revised items (see Appendix Table A1 for the trajectory of the scales and item modifications; Glynn *et al.*, 2009, 2011). Though this level of development and testing is commendable, it is a common misconception that once an instrument is published in the literature it is "validated" for all uses regardless of changes in population, context, or wording (Barbera and VandenPlas, 2011). The

proliferation of variations of the SMQ II in the STEM education literature provides opportunities to examine how frequently evidence is found to support the hypothesized five-factor structure of the SMQ II.

As the SMQ II developers intentionally designed the instrument such that the word ‘science’ could be replaced with any other specific discipline (Glynn *et al.*, 2011) many versions of the SMQ II can be found in the literature using wording such as biology, chemistry, organic chemistry, histology, math, nanotechnology, pharmacy, physics, and technology (Tosun, 2013; Campos-Sánchez *et al.*, 2014; Riccitelli, 2015; Salta and Koulougliotis, 2015; Srisawasdi, 2015; Hibbard *et al.*, 2016; Kassae, 2016; Kwon, 2016; Mahrou *et al.*, 2016; Olimpo *et al.*, 2016; Cleveland *et al.*, 2017; González *et al.*, 2017; Reece and Butler, 2017; Yamamura and Takehira, 2017; Ardura and Pérez-Bitrián, 2018; Austin *et al.*, 2018; Cagande and Jugar, 2018; Komperda, Hosbein, *et al.*, 2018; Young *et al.*, 2018). The popularity of the SMQ II also extends to discipline-based education researchers who have translated it from English into at least seven other languages (Tosun, 2013; Campos-Sánchez *et al.*, 2014; Salta and Koulougliotis, 2015; Srisawasdi, 2015; Schumm and Bogner, 2016; González *et al.*, 2017; Shin *et al.*, 2017; Yamamura and Takehira, 2017; Ardura and Pérez-Bitrián, 2018; Vasques *et al.*, 2018).

Investigation of the internal structure of the SMQ II has utilized analysis techniques both with and without *a priori* models of how the items should be related. Analyses without a prior model generally fall under the classification of exploratory factor analysis (EFA), although of the most commonly used techniques with the SMQ II, principal components analysis and principal axis factoring, the former is frequently described as a data reduction technique, not a factoring approach (Henson and Roberts, 2006). As described earlier, the theoretical framework of the SMQ II describes motivation as a “multicomponent construct” composed of “types and attributes



of motivation” (Glynn *et al.*, 2011, 1161) including intrinsic motivation, self-determination, self-efficacy, grade motivation, and career motivation. Of the seven studies using EFA techniques, four identified five factors aligned with the proposed theoretical framework (Glynn *et al.*, 2011; Kwon, 2016; Schmid and Bogner, 2017; Ardura and Pérez-Bitrián, 2018). After initially failing to find a five-factor solution, Austin *et al.*, (2018) removed a majority of the intrinsic items resulting in a combined intrinsic/career factor described as ‘relevance.’ Yamamura and Takehira (2017) also obtained a four-factor solution after removing 12 items due to low association with a factor, including all the self-efficacy items. The last study only utilized three scales from the SMQ II (self-efficacy, self-determination, and career) which resulted in a three-factor solution (Schumm and Bogner, 2016). These studies (Table A2) provide some support that the items are aligned with their intended factors but moving to a confirmatory framework provides the ability to test data against a previously specified model and restricts items to only associating with a single factor.

The SMQ II developers specified a correlated five-factor model with five items belonging to each factor (Glynn *et al.*, 2011). Therefore, data collected from administration of the SMQ II can be tested against this *a priori* model and evaluated with typical data-model fit criteria (Hu and Bentler, 1999). These data-model fit criteria take the form of examining the value of various fit indices and comparing them to suggested cutoff values, generally a CFI and/or TLI at or above 0.95, RMESA at or below 0.06, and SRMR at or below 0.08. Direct comparison of data-model fit across studies with the SMQ II is difficult due to variations in the wording of the items as either science or discipline-specific, and editing or removal of the items themselves (see Table A3 for a list of studies, variations, and fit values). However, in general the Hu and Bentler cutoff criteria were not met by most studies (Glynn *et al.*, 2011; Salta and Koulougliotis, 2015; Kwon,

2016; Ardura and Pérez-Bitrián, 2018; Komperda, Hosbein, *et al.*, 2018; Vasques *et al.*, 2018) unless the instrument was modified by removing items or entire scales (Tosun, 2013; González *et al.*, 2017; Yamamura and Takehira, 2017).

Additional limitations for the direct comparison of CFA results across studies are due to the low frequency with which information is reported about the estimator chosen for the factor analysis (Table A2) and justification that the properties of the data supported the use of the chosen estimator. For example, when descriptive statistics are reported for SMQ II items or scales it is frequently found that responses to the grade motivation items are much higher (more positive) than the other scales (Glynn *et al.*, 2011; Salta and Koulougliotis, 2015; Hibbard *et al.*, 2016; Ardura and Pérez-Bitrián, 2018; Austin *et al.*, 2018; Komperda, Hosbein, *et al.*, 2018). This could indicate potential issues with nonnormality of data or collapsing of the five-point response scale such that it essentially functions only as a two- or three-point scale for some items. In any of these cases it would be recommended to move from the typical maximum likelihood (ML) estimator to a robust estimator (MLR) that provides a correction for non-normality or a mean and variance adjusted weighted least squares (WLSMV) estimator for categorical data (Finney & DiStefano, 2013). Studies employing the WLSMV estimator with the full 25-item SMQ II have found slightly better data-model fit than those employing the ML estimator (Komperda, Hosbein, *et al.*, 2018). The inconsistency among the CFA results suggests the need to examine causes for this variation, particularly as they related to alignment between the theoretical framework of the SMQ II, the individual items, and student responses. Providing evidence of these alignments is paramount to ensuring that the instrument has solid theoretical support and can further be used to inform instructional practices.

## Research goals

In light of inconsistent evidence for the internal structure of the SMQ II in the literature, this research investigated modifications that could potentially improve the functioning of the instrument, as measured within a factor analysis framework. Prior work by Komperda, Hosbein, and Barbera with the original 25-item version of the SMQ II (2018) had identified both overall poor data-model fit and differences in data-model fit across wordings (science and chemistry) and course types (introductory chemistry and general chemistry) but did not explore possible explanations for the model misfit or alternative items to improve model fit. In order for instructors or other researchers to be able to utilize the SMQ II for comparisons of motivational impacts, the instrument needs to show similar psychometric characteristics across the varied measurement contexts. Therefore, the overall goal of the current research was to determine if modifications could be made in such a way that both improved overall data-model fit and minimized differences in data-model fit across different measurement contexts, that is, in different types of courses with both the science and discipline-specific wording. These outcomes would align with the original goals of the SMQ II developers to have evidence for the functioning of discipline-specific versions of the instrument. This work was driven by two primary research questions.

1. What are potential reasons for the inconsistent validity evidence based on the internal structure of the SMQ II as proposed by the SMQ II developers?
2. If these issues are addressed, will a modified SMQ II that is aligned with the theoretical framework proposed by the SMQ II developers have acceptable internal structure across different wordings and course contexts?

These research questions were addressed in two phases with independent samples. The goal of phase one was to identify modifications that could be made to the SMQ II to improve the functioning of both the science and discipline-specific wordings when administered to undergraduate students. The result of phase one was the development of a modified Science Motivation Questionnaire II (mSMQ II). The goal of phase two was to assess the functioning of the mSMQ II in a new sample of undergraduate students enrolled in science courses in order to evaluate whether the modifications resulted in improved data-model fit relative to the SMQ II. A factor analysis framework was chosen for this research to align with previous work done by the instrument developers (Glynn *et al.*, 2011) and to provide a point of comparison to the previously discussed SMQ II studies. The methods and results from each phase are reported sequentially. Within each phase, human subjects IRB approval was obtained from Portland State University and appropriate participant consent was gathered from the study populations. Any incentives, if provided, are noted within each population description.

## **Phase one**

### **Methods**

In light of the limited evidence provided in the literature for the structural validity of SMQ II scores, the first phase of this research investigated potential threats to validity due to interpretation of the response scale and individual items. As the internal structure of an instrument can be impacted by issues with the response scale and/or the interpretation of items by the target population, three sources of information were utilized in phase one: expert reviews, student response processes, and best practices in survey wording. Expert reviews were solicited using an online survey of experts in educational research and measurement, asking them to determine if the five-point frequency-based response scale on the original SMQ II, ranging from

never to always, was appropriately aligned with the wording of the items themselves. Interviews on a subset of SMQ II items were conducted with a convenience sample of general chemistry students to better understand their response processes related to both the frequency-based response scale and their interpretation of item wording. The research team used the results of the expert response scale survey, the student interviews, and best practices in survey wording (Krosnick and Presser, 2010) to make modifications to the SMQ II, resulting in the mSMQ II.

### ***Participants and data collection***

#### ***Expert response scale survey***

A total of 12 experts with experience in discipline-based education research (n=8) or educational measurement (n=4) were invited to and participated in an online survey about the type of response scale that was best suited to the wording of each SMQ II item. Experts were blinded to the fact that the items were from the SMQ II and that the original response scale was frequency-based. Experts were asked to sort items based on their perception of whether each item could best be answered with a frequency-based scale (never-rarely-sometimes-usually-always), a Likert-type scale (disagree-somewhat disagree-neither disagree nor agree-somewhat agree-agree), or that both scales would be equally appropriate. The survey was conducted in the same term as the student interviews and the experts were not compensated for their time.

#### ***Student interviews***

Student interviews on the SMQ II items were conducted as part of a larger project investigating how wording of survey items affects student responses. Students were recruited during the winter 2017 term from both on-sequence and off-sequence large-enrollment general chemistry courses for science majors at Portland State University. A total of 40 student interviews were conducted, representing 5% of the overall course enrollment. The students

participating in the interviews represent only a subset of students ultimately involved in the larger psychometric analyses conducted in phase 2 of this project. However, these students are representative of the population in which the SMQ II has been utilized, which is with university-level students in their first or second year of a majors-level science curriculum.

During the interview, students were provided a paper copy of the SMQ II and randomly presented with either the science or chemistry wording of the items. Students read all of the items silently and circled their responses on the original frequency-based scale. Students were asked to explain their responses to a subset of items, which were identified by the research team as having potentially good or poor fits to the response scale and or a hypothesized item category based on a prior study (Komperda, Hosbein, *et al.*, 2018). Next, students were asked to go over the entire survey again and explain if any item responses would change if the other wording (science or chemistry) were substituted. Specific demographics were not collected for the interview participants and each was provided a \$10 gift card for their time.

## ***Analysis***

### ***Expert response scale survey analysis***

Responses to the expert survey were summarized as percentages to identify which items experts perceived to be more aligned with the original five-point frequency-based response scale (never-rarely-sometimes-usually-always), a five-point Likert-type (disagree-somewhat disagree-neither disagree nor agree-somewhat agree-agree) response scale, or where the response scales were seen as equally appropriate for the wording of the item.

### ***Student interview coding***

The audio files of the interviews were transcribed by a commercial transcription service. All analysis was conducted on the interview transcripts, led by author Hosbein. Cognitive

interview analysis (Beatty, 2004) was used by author Hosbein to establish a coding rubric. This entailed carefully reading each transcript and documenting the language students used in explaining their responses. These reviews were used to establish initial codes regarding students' scale-based language. Example codes are provided in Table A5. All coding was completed using QSR NVivo 8. Author Hosbein and a secondary (non-author) researcher independently coded all 40 transcripts using the initial rubric. After discussing discrepancies in coding, the rubric was revised and additional transcripts were independently re-coded. The final rubric had an acceptable Cohen's Kappa value of 0.8 (Hallgren, 2012) and was used by author Hosbein to re-code the remaining transcripts.

## **Results and discussion**

### **Expert response scale survey**

Expert preference for a response scale was determined by simple majority for one of the two response scales. Experts showed a preference for the original frequency-based response scale for only five of the 25 items, whereas for 12 items experts preferred the Likert-type response scale. For the remaining items, experts either showed no majority preference for either response scale, or felt that either scale would be acceptable. Overall, self-determination labeled items were most likely to be judged to fit best on the frequency-based scale whereas items labeled as intrinsic, self-efficacy, grade, and career motivation were more likely to be judged to align with a Likert-type scale. Detailed results are provided in Table A4.

### **Student interviews**

Student responses were coded as frequency-based when a student explicitly used the frequency-based scale words (never, rarely, sometimes, usually, always) or other time-based words (often, typically, never really, and most of the time) in their response. Students also used

the frequency scale in more of a quantity-based way with responses more aligned with “how much” rather than “how often.” A quantity code was assigned when a student used language that involved quantity in their response that was not specifically frequency-based or made a comparison between two things that implied a quantity. Language that reflected the quantity code included the words, “really”, “a lot”, or “it depends.” If both frequency-based and quantity-based language was used to explain a response, the response was coded as frequency-based. Examples of each code are provided in the Table A5.

Though students were not asked to fully describe their response process for all SMQ II items during the interviews, the results from the 12 items students responded to show similarities between language used by students and the preferred response scale identified by the experts. The three intrinsic labeled items explored in the interviews (I2, I3, and I5; wording given in Table 1) were coded as having frequency-based responses in less than 40% of instances (32%, 25%, and 38%, respectively), which aligned with expert preferences not to use a frequency-based response scale for these items. Similar responses were seen for the self-efficacy item SE2 in which frequency-based codes were used with 32% of responses and for the three career items explored in the interviews (C1, C2, and C3) with less than a quarter of students using frequency-based language (15%, 22%, and 15%, respectively). Only two items explored in the interviews, a self-efficacy item (SE1) and a grade item (G4) showed a majority use of frequency-based language (65% and 57% of codes, respectively), which is also consistent with the experts’ evaluation of being more aligned with a frequency-based response scale than a Likert-type scale.

When asking students to explain their responses to some of the SMQ II items, recurrent issues with students' interpretation arose. For example, when explaining their responses to the intrinsic item (I1) "The science I learn is relevant to my life", 17% of students referenced their



career as the reason that science (or chemistry) was relevant to their lives. This was unexpected since the SMQ II contains a separate set of items intended to address students' career motivation. Similar overlap with thinking about future careers was seen in responses to the grade item (G4) "I think about the grade I will get in science" students cited pressure from graduate or professional school for the reason they think about their grade.

When responding to the intrinsic item (I3) "Learning science makes my life more meaningful," students were unsure of what "meaningful" meant in that context. The most frequent way students described "meaningful" was that learning science (or chemistry) helped them to better understand the world around them. Students also expressed confusion about other vague phrases such as "relevant" (I1) and "think about" (G4) suggesting that the wording of these items could be made clearer to improve response process validity.

Another commonly observed response was for students to ignore a portion of an item when formulating their response. An example of this is with the self-efficacy item (SE2) "I am confident I will do well on science labs and projects." Some students explicitly mentioned only focusing on the lab portion of the question because their course did not involve projects while other students responding to this question made a comparison between labs and tests. In these instances, students may be interpreting projects to mean tests or simply ignoring the project portion of the question. In either case, this suggests problems with the wording of the item due to the presence of multiple topics within a single item (i.e., the item is double-barreled) or a topic in the item not being applicable to the typical experience of a student in a general chemistry course (e.g., having projects). Full student quotes are provided in the Appendix, Table A6.

## ***Revisions to SMQ II***

The results of both the expert response scale surveys and student interviews were in general agreement that items on the grade, career, and intrinsic scales were not well aligned with a frequency-based response scale. In lieu of using two different response scales within the same instrument, the modified SMQ II (mSMQ II) response scale was changed to the more traditional Likert-type response scale containing five scale points (disagree-somewhat disagree-neither disagree nor agree-somewhat agree-agree), as presented in the expert response scale survey. Similar changes have been made by other users of the SMQ II (Srisawasdi, 2015; Kwon, 2016; Olimpo *et al.*, 2016; Childers and Jones, 2017; Schmid and Bogner, 2017).

During the student interviews, two intrinsic items (I1 and I3) were identified as potentially causing unintended student responses. In the first item, I1, students were considering their career as a reason that science or chemistry was relevant to their lives, which could cause this item to be more associated with responses to items on the career scale rather than the intrinsic motivation scale. For that reason, an additional version of the item “The science I learn is relevant to the world around me” was added to the survey in order to test whether more general wording could avoid having the item align with the career items. The second item, I3, was revised into two different wordings to address different reasons for why science or a specific discipline may be “meaningful” to students, which is by increasing their understanding or appreciation. The phrase “world around me” was used in these revisions to align with the revision to item I1. Additionally, students describing their response to item G4 emphasized thinking about their grade in terms of worrying about requirements for graduate or professional school so an additional item was added to explore the “worry” aspect of thinking about grades.

Other modifications were made to items to remove or separate double-barreled items. As in Salta and Koulougliotis (2015), the phrase “labs” was removed from items SD4, SE2, and G5 rather than removing the items entirely (Austin *et al.*, 2018) since most of the classes had separate lab and lecture components. Additionally, the phrase “projects” was removed from SE2 and replaced with the more general “assignments” since not all classes have projects. Similarly, item SE3 had “skills” removed since that appeared more aligned with a laboratory context.

The final set of modifications was made to deal with phrases that were either overly vague, or overly focused on a specific aspect of course performance. One self-determination item, SD1, “I put enough effort into learning science” was modified into the more concrete “I put effort into learning science well”, which better aligned it with item SD2 about using “strategies to learn science well.” Similarly, a second version of item SD5 “I study hard to learn science” was written more concretely as “I use a lot of mental energy learning science.” In a grade item, G1, specifically focused on students doing better than others on tests, the test-specific language was dropped to account for other aspects of course performance. Similarly, two items, SE4 and G3, focused students on earning an “A” in science, in recognition that this is not necessarily the goal for all students, this was reworded as earning “the grade I want in science.” The complete set of mSMQ II items is provided in Table 1. Items that stayed the same from the SMQ II to the mSMQ II retain the same numbering. Items that have been revised have a letter after their designation (e.g., I1a) to indicate that they are a revision of a pre-existing SMQ II item, I1 in this case. In total, 29 items appeared on the mSMQ II.

**Table 1.** Scale, label, and wording for SMQ II and mSMQ II items where [ ] was replaced with either biology, chemistry, or science.

Scale	Item	SMQ II Wording	<sup>a</sup> Item	mSMQ II Wording
Intrinsic	I1	The science I learn is relevant to my life	I1*	The [ ] I learn is relevant to my life
	I2	Learning science is interesting	I1a*	The [ ] I learn is relevant to the world around me
	I3	Learning science makes my life more meaningful	I2	Learning [ ] is interesting
	I4	I am curious about discoveries in science	I3a	Learning [ ] helps me understand the world around me
	I5	I enjoy learning science	I3b	Learning [ ] increases my appreciation of the world around me
Self-determination	SD1	I put enough effort into learning science	I4	I am curious about discoveries in [ ]
	SD2	I use strategies to learn science well	I5	I enjoy learning [ ]
	SD3	I spend a lot of time learning science	SD1a	I put effort into learning [ ] well
	SD4	I prepare well for science tests and labs	SD2*	I use strategies to learn [ ] well
	SD5	I study hard to learn science	SD3	I spend a lot of time learning [ ]
Self-efficacy	SE1	I am confident I will do well on science tests	SD4a*	I prepare well for [ ] tests
	SE2	I am confident I will do well on science labs and projects	SD5	I study hard to learn [ ]
	SE3	I believe I can master science knowledge and skills	SD5a*	I use a lot of mental energy learning [ ]
	SE4	I believe I can earn a grade of “A” in science	SE1	I am confident I will do well on [ ] tests
	SE5	I am sure I can understand science	SE2a	I am confident I will do well on [ ] assignments
Grade	G1	I like to do better than the other students on science tests	SE3a*	I believe I can master [ ] knowledge
	G2	Getting a good science grade is important to me	SE4a	I believe I can earn the grade I want in [ ]
	G3	It is important that get an “A” in science	SE5*	I am sure I can understand [ ]
	G4	I think about the grade I will get in science	G1a*	I like to do better than the other students in [ ]
	G5	Scoring high on science tests and labs matters to me	G2	Getting a good [ ] grade is important to me
Career	C1	Learning science will help me get a good job	G3a	It is important that I earn the grade I want in [ ]
	C2	Knowing science will give me a career advantage	G4	I think about the grade I will get in [ ]
	C3	Understanding science will benefit me in my career	G4a*	I worry about my [ ] grade
	C4	My career will involve science	G5a	Scoring high on [ ] tests matters to me
	C5	I will use science problem-solving skills in my career	C1	Learning [ ] will help me get a good job
			C2	Knowing [ ] will give me a career advantage
			C3	Understanding [ ] will benefit me in my career
			C4	My career will involve [ ]
			C5*	I will use [ ] problem-solving skills in my career

<sup>a</sup>Items removed after exploratory factor analysis are indicated with asterisks.

## Phase two

### Methods

In phase two, the mSMQ II was administered to a nationwide sample of undergraduate students using both the science and discipline-specific wordings, which were aligned with the discipline of the course in which the students were enrolled, either biology or chemistry. Following a similar process to that used by the SMQ II authors (Glynn *et al.*, 2011), the mSMQ II responses were randomly split into two equally sized datasets balanced across course, wording, and self-reported gender of respondent to avoid unintentional bias in the datasets. Exploratory factor analysis (EFA) was used on the first half of the data (the ‘training’ dataset) to examine potential issues with the functioning of the modified and added items. Results from the EFA were used to identify and remove problematic items. Full details of the EFA methods and results are contained in the Appendix. Confirmatory factor analysis (CFA) was used with the remaining data (the ‘testing’ dataset) to test the data-model fit for a five-factor model of the mSMQ II with the reduced set of items, representing the structure hypothesized by the original developers. Additional CFAs were performed on single-factor models representing the individual aspects of motivation comprising the mSMQ II. These individual factor models were also used to provide information about the reliability of the individual motivation scales.

### *Participants and data collection*

Student participants were recruited by contacting instructors of chemistry and biology courses at nine different colleges and universities across the United States. Courses were classified as preparatory chemistry if the official description indicated that the course was designed to prepare academically weaker students for eventual enrollment in a general chemistry course. Courses were classified as general chemistry if the official course description either had a

required or recommended prerequisite of secondary school chemistry or equivalent or if the description indicated that the course was designed for science or engineering majors. Courses were classified as general biology if the official course description indicated that the course was designed for science majors. Data were collected in both on-sequence first term and off-sequence second term general chemistry courses and general biology courses, with data for each discipline combined into one dataset.

A link to the online mSMQ II, created in Qualtrics, was provided to each course instructor. The instructor was asked to provide this link to students through their course management website and also to play a brief video in class in which a research team member described the purpose of the study and the consent process to students. No identifying student information was collected on the survey itself. Most of the course instructors offered extra credit for student participation in the survey. If extra credit was offered, students were taken to a separate survey where they entered their name and university ID for identification purposes for extra credit only. All surveys were open for a non-exam week selected by the instructor between the end of October and the end of November 2017. When taking the survey, students were randomly presented with either the science or discipline-specific wording (biology or chemistry) for all mSMQ II items. The items were presented in a randomized order followed by demographic questions about gender, race/ethnicity, and declared major.

### *Analysis*

All data cleaning and analysis steps for phase two were performed using R version 3.5.0 (R Core Team, 2019).

### ***Data cleaning***

A total of 3,386 raw survey responses were obtained. Responses for an entire course were excluded from the full data set if the course response rate was under 25% of enrollment, leaving 3,101 responses. Next, individual responses were removed if the student did not correctly answer a ‘participant check’ item asking students to select “Disagree.” Additionally, missing data were addressed through listwise deletion. As a result of these cleaning steps the final data set contained 2,487 responses corresponding to 73% of the raw data.

### ***Response patterns and descriptive statistics***

Descriptive statistics including mean, median, standard deviation, range, skew and kurtosis were calculated for each wording and course condition using the R psych package (version 1.8.4; Revelle, 2018).

### ***Confirmatory factor analysis***

After identifying and removing problematic items based on the EFA results (see EFA section in Appendix), confirmatory factor analysis (CFA) was used to test models of the mSMQ II with the reduced number of items (as noted in Table 1). In line with the original model used by the developers, and supported by the EFA results, a correlated five-factor model of the mSMQ II items was tested for all course and wording conditions using the testing dataset previously partitioned from the general biology and general chemistry courses and the full preparatory chemistry data set. A training dataset was not generated from the preparatory chemistry data due to the smaller sample size. Lastly, using the combined training and testing data sets, single-factor models were tested for each of the five reduced sets of items representing aspects of motivation (e.g., intrinsic, career, etc.).

The self-determination and self-efficacy scales of the mSMQ II only consisted of three items each after removing poorly functioning items. With only three items and no restrictions on the strength of associations between an item and a factor, known as loadings, a single-factor model has zero degrees of freedom and data-model fit cannot be tested. Constraining loadings on a factor to be equal (i.e., a tau-equivalent model) restores degrees of freedom to the model and data-model fit can be tested (Komperda, Pentecost, *et al.*, 2018). While tau-equivalent models are more restrictive and therefore less likely to achieve model-fit than unconstrained (i.e., congeneric) models, it is necessary to use them when a factor has less than four items for the aforementioned reasons. Therefore, tau-equivalent single-factor models were tested for the self-determination and self-efficacy scales, while congeneric single-factor models were tested for the intrinsic, grade, and career scales.

In recognition of the ordinal and highly-skewed properties of the mSMQ II data, the robust diagonally weighted least squares (WLSMV) estimator was used (Finney and DiStefano, 2013). As the WLSMV estimator was expected to show better data-model fit than robust maximum likelihood (MLR) due to the properties of the data; fit indices from both estimators are provided for comparison purposes. Data-model fit was evaluated using a set of indices appropriate for the estimator used (Hu and Bentler, 1999; Yu, 2002; Beauducel and Herzberg, 2006; Xia and Yang, 2018). For the WLSMV estimator, values of CFI and TLI  $\geq 0.95$  and RMSEA  $\leq 0.05$  were used to indicate acceptable data-model fit. Since previous studies demonstrated that the SRMR does not function well with the WLSMV estimator with small number of response categories, the SRMR was not used to make data-model fit assessments for this estimator. For the MLR estimator, values of CFI and TLI  $\geq 0.95$ , RMSEA  $\leq 0.06$ , and SRMR  $\leq 0.08$  were used to determine acceptable data-model fit. For both estimators, a model



was deemed to have acceptable data-model fit when all fit indices were acceptable. All CFA models were analyzed using the lavaan package in R (version 0.6–1; Rosseel, 2012).

### ***Reliability***

There are multiple ways to assess the reliability of data obtained from a survey instrument (American Educational Research Association *et al.*, 2014); for this research single-administration reliability values were reported since the instrument was only administered at one time point and the sample size was large enough to examine the internal structure of the instrument (Komperda, Pentecost, *et al.*, 2018). It is not appropriate to address the reliability of the instrument as a whole due to the multidimensional way in which motivation was conceptualized by the original developers as incorporating five distinct factors (Cronbach, 1951). Instead, single-administration reliability values were calculated for each motivation scale showing acceptable data-model fit to a single-factor model tested during the CFA portion of the analysis. All reliability calculations were performed using polychoric correlations to account for the ordinal nature of the data (Gadermann *et al.*, 2012) with the userfriendlyscience R package (version 0.7.1; Peters, 2017).

## **Results and discussion**

### ***Participant characteristics***

After data cleaning there were 2,487 usable responses from students who responded to all items on the survey. A majority of the students surveyed were enrolled in general chemistry courses (Table 2). Across the various course and wording conditions students were primarily female (ranging from 51-77%) and white (ranging from 60-75%). Students enrolled in the preparatory chemistry courses and general biology courses were primarily biology pre-health majors. In general chemistry, students were primarily engineering majors (30% of science

wording and 31% of chemistry wording) or biology pre-health majors (21% of science wording and 20% of chemistry wording).

**Table 2.** Student responses and demographic information for mSMQ II administration.

Course	Wording	Responses	Female	White	Top major (%)
Preparatory	Science	139	61%	75%	Biology pre-health (40%)
Chemistry	Chemistry	137	76%	74%	Biology pre-health (35%)
General	Science	835	55%	63%	Engineering (30%)
Chemistry	Chemistry	855	51%	60%	Engineering (31%)
General	Science	258	77%	61%	Biology pre-health (29%)
Biology	Biology	263	76%	66%	Biology pre-health (28%)

### *Descriptive statistics*

Responses to the mSMQ II items followed a similar pattern to previous studies (Komperda, Hosbein, *et al.*, 2018) where students were more likely to respond positively (4 or 5 on the 5-point response scale) to items with the science wording relative to the discipline-specific wording. These differential response patterns were most pronounced for items associated with intrinsic and career motivation. Of the five motivation aspects, the grade motivation items had the most strongly positive responses, regardless of wording, which is aligned with what other researchers have reported (Glynn *et al.*, 2011; Salta and Koulougliotis, 2015; Hibbard *et al.*, 2016; Ardura and Pérez-Bitrián, 2018; Austin *et al.*, 2018; Komperda, Hosbein, *et al.*, 2018). Detailed descriptive statistics are provided in Table A7.

### *Exploratory factor analysis*

The EFA results were used to identify potentially problematic items that should be removed before moving into a confirmatory framework. Items were determined to be problematic if they showed low relation to their intended scale factor, if they showed evidence of association with more than one factor, or if they displayed an inconsistent pattern of association with a factor across different wording and course conditions. This last condition is particularly important for the mSMQ II since the original SMQ II is intended to be used to measure

motivation in different contexts. Items identified as problematic are indicated in Table 1 and Figure A1, with an asterisks below their coefficient bar for the factor they were intended to be associated with. The full results of the EFA can be found in the Appendix.

### ***Confirmatory factor analysis***

#### ***Five-factor models***

The correlated five-factor models of the mSMQ II data represent the hypothesized framework of motivation proposed by the original developers. Testing the fit between this hypothesized model and the data collected from the mSMQ II provides evidence for the validity of that underlying theoretical framework. As expected, since the WLSMV estimator was more appropriate for the characteristics of the data, the fit indices for the 19 mSMQ II items in the five-factor model with the WLSMV estimator reached acceptable levels for more course and wording combinations than the MLR estimator (see Table 3). Overall these fit indices were better than previous studies using the WLSMV estimator (Komperda, Hosbein, *et al.*, 2018) and more consistently acceptable across course and wording combinations suggesting that the extensive revisions minimized some previous issues with instrument functioning across course and wording combinations. For WLSMV, only the general chemistry course with the chemistry wording failed to meet acceptable cutoff values whereas with MLR half of the course and wording combinations failed to meet acceptable data-model fit (both wordings in preparatory chemistry and the science wording in general biology).

The acceptable fit index values for some wording and course combinations with the MLR estimator align with those seen in other studies involving extensive modifications by removing items (Tosun, 2013; González *et al.*, 2017; Yamamura and Takehira, 2017). Unfortunately, it is unclear if those studies with acceptable fit indices used the same estimator, though it is a

reasonable assumption since the default estimator in many CFA programs is maximum likelihood. The MLR fit indices that did not meet acceptable values are more similar to those from studies by the original developers (CFI = 0.91; RMSEA = 0.07; SRMR = 0.04; Glynn *et al.*, 2011) or those using the 25 SMQ II items with only modifications to the language (e.g., Greek) or the target (e.g., chemistry) of the instrument (Salta and Koulougliotis, 2015; Kwon, 2016; Ardura and Pérez-Bitrián, 2018; Vasques *et al.*, 2018).

The results of testing the five-factor model with two different estimators suggests two possibilities. First, that the poor data-model fit seen in prior work with the unmodified SMQ II was primarily a result of using an inappropriate estimator for the characteristics of the data. However, this conclusion is suspect because previous work using the original wording and the appropriate WLSMV estimator did not show consistently acceptable data-model fit (Komperda, Hosbein, *et al.*, 2018). Second, that the need for extensive revisions or removal of items in order to fit a five-factor model, as in this study and others, indicates a larger problem with the underlying theoretical framework of the instrument. This second possibility was investigated by looking at how the individual scales, which represent the individual aspects of motivation function as independent factors. If the scales show good data-model fit as single-factor scales this indicates that they are appropriate measurements of a construct but that they do not necessarily relate to each other in the ways hypothesized by the SMQ II developers. If the individual scales do not show good data-model fit as single-factor models this indicates issues with what the scales themselves are measuring and whether it is well aligned with existing theories of motivation.

**Table 3.** Data-model fit for five-factor mSMQ II model (df = 142) with WLSMV and MLR estimators. Acceptable individual data-model fit indices are noted in bold. A model is deemed ‘acceptable’ when all indices for a given course/wording data set are bolded.

Estimator	Course	Wording	$\chi^2$	<sup>a</sup> CFI	<sup>a</sup> TLI	<sup>a</sup> RMSEA	[90% CI]	<sup>a</sup> SRMR
WLSMV	Preparatory	<b>Science (n=139)</b>	189	<b>0.99</b>	<b>0.99</b>	<b>0.05</b>	[0.03, 0.07]	-
	Chemistry	<b>Chemistry (n=137)</b>	194	<b>0.99</b>	<b>0.98</b>	<b>0.05</b>	[0.03, 0.07]	-
	General	<b>Science (n=417)</b>	251	<b>0.99</b>	<b>0.98</b>	<b>0.04</b>	[0.03, 0.05]	-
	Chemistry	Chemistry (n=426)	334	<b>0.99</b>	<b>0.98</b>	0.06	[0.05, 0.06]	-
	General	<b>Science (n=128)</b>	179	<b>0.99</b>	<b>0.98</b>	<b>0.05</b>	[0.02, 0.06]	-
	Biology	<b>Biology (n=130)</b>	191	<b>0.99</b>	<b>0.99</b>	<b>0.05</b>	[0.03, 0.07]	-
MLR	Preparatory	Science (n=139)	251	0.90	0.88	0.07	[0.06, 0.09]	<b>0.06</b>
	Chemistry	Chemistry (n=137)	235	0.93	0.92	0.07	[0.05, 0.08]	<b>0.06</b>
	General	<b>Science (n=417)</b>	227	<b>0.97</b>	<b>0.96</b>	<b>0.04</b>	<b>[0.03, 0.05]</b>	<b>0.04</b>
	Chemistry	<b>Chemistry (n=426)</b>	284	<b>0.96</b>	<b>0.95</b>	<b>0.05</b>	<b>[0.04, 0.06]</b>	<b>0.04</b>
	General	Science (n=128)	255	0.89	0.87	0.08	[0.06, 0.09]	<b>0.07</b>
	Biology	<b>Biology (n=130)</b>	210	<b>0.96</b>	<b>0.95</b>	<b>0.06</b>	<b>[0.04, 0.08]</b>	<b>0.05</b>

<sup>a</sup>Acceptable data-model fit values differ by estimator: for WLSMV cut-off values are CFI and TLI  $\geq 0.95$  and RMSEA  $\leq 0.05$ ; for MLR cut-off values are CFI and TLI  $\geq 0.95$ , RMSEA  $\leq 0.06$ , and SRMR  $\leq 0.08$ .

### *Single-factor models*

Unlike in the five-factor model results, there is less of a clear benefit from using the WLSMV estimator relative to MLR in the single-factor models. Of the 30 possible course and wording combinations, only seven had acceptable data-model fit under the WLSMV estimator whereas 11 were acceptable using MLR (shown in bold in Table 4). None of the scales showed consistently acceptable data-model fit across all course and wording combinations. Unexpectedly, the models for self-determination and self-efficacy with the loadings constrained to be equal across items were as likely to have acceptable data-model fit as other scales where this constraint was not present.

Examination of patterns of acceptable and unacceptable data-model fit for the single factor models provides evidence that the model of the remaining intrinsic items (I2, I3a, I3b, I4, and I5) showed consistently poor data-model fit across course and wording conditions, particularly in the high RMSEA values relative to other scales. In contrast, the grade motivation items had a majority of course and wording combinations with acceptable data-model fit, more than any other scale. The patterns are more difficult to interpret for the self-determination and

self-efficacy items given their inconsistency in fit across course and wording combinations as well as the additional constraints placed on those scales in order to obtain data-model fit information. Similarly, the results for the career items were somewhat mixed particularly when comparing across estimators.

Overall, the single-factor model results provide some evidence that the underlying issue of inconsistent fit of the SMQ II and mSMQ II data may be due to problems with the individual aspects of motivation hypothesized to underpin the instrument. Single-factor models of the original SMQ II items showed similar patterns with the WLSMV estimator in that the intrinsic items had less support for their structure across course and wording combinations while the grade motivation items had more (Komperda, Hosbein, *et al.*, 2018). Yet, grade motivation is not a known construct within the motivation literature while intrinsic motivation is. These results further support the interpretation that even when a more appropriate estimator is used there are underlying issues in the structure and framework of the SMQ II and mSMQ II responsible for the poor data-model fit.

**Table 4.** Data-model fit for single-factor mSMQ II scales with WLSMV and MLR estimators. Acceptable individual data-model fit indices are noted in bold. A model is deemed ‘acceptable’ when all indices for a given course/wording data set are bolded.

Model	Estimator	Course	Wording	$\chi^2$	<sup>a</sup> CFI	<sup>a</sup> TLI	<sup>a</sup> RMSEA	[90% CI]	<sup>a</sup> SRMR
Intrinsic items ( <i>df</i> = 5)	WLSMV	Preparatory	Science (n=139)	18	<b>0.99</b>	<b>0.98</b>	0.14	[0.08, 0.21]	-
		Chemistry	Chemistry (n=137)	42	<b>0.98</b>	<b>0.96</b>	0.23	[0.17, 0.30]	-
		General	Science (n=835)	69	<b>0.99</b>	<b>0.98</b>	0.12	[0.10, 0.15]	-
		Chemistry	Chemistry (n=855)	166	<b>0.98</b>	<b>0.96</b>	0.19	[0.17, 0.22]	-
		General	Science (n=258)	37	<b>0.98</b>	<b>0.97</b>	0.16	[0.11, 0.21]	-
	Biology	Biology (n=263)	48	<b>0.99</b>	<b>0.97</b>	0.18	[0.14, 0.23]	-	
	MLR	Preparatory	Science (n=139)	18	0.93	0.86	0.14	[0.09, 0.20]	<b>0.04</b>
		Chemistry	Chemistry (n=137)	34	0.91	0.81	0.21	[0.15, 0.27]	<b>0.04</b>
		General	Science (n=835)	31	<b>0.97</b>	<b>0.95</b>	0.08	[0.06, 0.10]	<b>0.03</b>
		Chemistry	Chemistry (n=855)	107	0.94	0.87	0.15	[0.13, 0.18]	<b>0.04</b>
General		Science (n=258)	44	0.90	0.81	0.17	[0.14, 0.22]	<b>0.05</b>	
Biology	Biology (n=263)	50	0.92	0.83	0.18	[0.15, 0.23]	<b>0.04</b>		
Self-determination items (equal loadings; <i>df</i> = 2)	WLSMV	Preparatory	<b>Science (n=139)</b>	0	<b>1.00</b>	<b>1.01</b>	<b>0.00</b>	<b>[0.00, 0.06]</b>	-
		Chemistry	Chemistry (n=137)	7	<b>0.98</b>	<b>0.97</b>	0.14	[0.04, 0.25]	-
		General	Science (n=835)	26	<b>0.99</b>	<b>0.98</b>	0.12	[0.08, 0.16]	-
		Chemistry	Chemistry (n=855)	26	<b>0.99</b>	<b>0.98</b>	0.12	[0.08, 0.16]	-
		General	<b>Science (n=258)</b>	2	<b>1.00</b>	<b>1.00</b>	<b>0.02</b>	[0.00, 0.13]	-
	Biology	Biology (n=263)	11	<b>0.99</b>	<b>0.99</b>	0.13	[0.06, 0.21]	-	
	MLR	Preparatory	Science (n=139)	1	<b>1.00</b>	<b>1.02</b>	<b>0.00</b>	[0.00, 0.12]	0.09
		Chemistry	Chemistry (n=137)	10	0.89	0.84	0.17	[0.08, 0.28]	0.15
		General	Science (n=835)	13	<b>0.97</b>	<b>0.95</b>	0.08	[0.05, 0.12]	0.12
		Chemistry	Chemistry (n=855)	12	<b>0.98</b>	<b>0.97</b>	0.08	[0.05, 0.11]	0.09
General		<b>Science (n=258)</b>	0	<b>1.00</b>	<b>1.02</b>	<b>0.00</b>	<b>[0.00, 0.04]</b>	<b>0.02</b>	
Biology	<b>Biology (n=263)</b>	0	<b>1.00</b>	<b>1.02</b>	<b>0.00</b>	<b>[0.00, 0.04]</b>	<b>0.02</b>		
Self-efficacy items (equal loadings; <i>df</i> = 2)	WLSMV	Preparatory	<b>Science (n=139)</b>	2	<b>1.00</b>	<b>1.00</b>	<b>0.00</b>	[0.00, 0.16]	-
		Chemistry	Chemistry (n=137)	5	<b>1.00</b>	<b>1.00</b>	0.10	[0.00, 0.22]	-
		General	Science (n=835)	18	<b>1.00</b>	<b>0.99</b>	0.10	[0.06, 0.14]	-
		Chemistry	Chemistry (n=855)	10	<b>1.00</b>	<b>1.00</b>	0.07	[0.03, 0.11]	-
		General	Science (n=258)	13	<b>0.99</b>	<b>0.99</b>	0.14	[0.07, 0.22]	-
	Biology	Biology (n=263)	10	<b>0.99</b>	<b>0.99</b>	0.13	[0.06, 0.21]	-	
	MLR	Preparatory	Science (n=139)	3	<b>0.99</b>	<b>0.98</b>	<b>0.06</b>	[0.00, 0.16]	0.10
		Chemistry	Chemistry (n=137)	6	<b>0.97</b>	<b>0.96</b>	0.12	[0.00, 0.24]	0.09
		General	Science (n=835)	24	<b>0.97</b>	<b>0.95</b>	0.12	[0.08, 0.15]	0.09
		Chemistry	<b>Chemistry (n=855)</b>	9	<b>0.99</b>	<b>0.98</b>	<b>0.06</b>	[0.03, 0.11]	<b>0.05</b>
General		Science (n=258)	10	<b>0.96</b>	0.94	0.13	[0.07, 0.19]	0.13	
Biology	Biology (n=263)	7	<b>0.97</b>	<b>0.96</b>	0.10	[0.03, 0.19]	<b>0.08</b>		

Model	Estimator	Course	Wording	$\chi^2$	<sup>a</sup> CFI	<sup>a</sup> TLI	<sup>a</sup> RMSEA	[90% CI]	<sup>a</sup> SRMR		
Grade items ( <i>df</i> = 2)	WLSMV	Preparatory	Science (n=139)	7	<b>0.99</b>	<b>0.97</b>	0.14	[0.04, 0.25]	-		
		Chemistry	Chemistry (n=137)	3	<b>1.00</b>	<b>0.99</b>	0.06	[0.00, 0.19]	-		
		General	<b>Science (n=835)</b>	4	<b>1.00</b>	<b>1.00</b>	<b>0.04</b>	[0.00, 0.09]	-		
		Chemistry	Chemistry (n=855)	7	<b>1.00</b>	<b>1.00</b>	0.06	[0.02, 0.10]	-		
		General	<b>Science (n=258)</b>	2	<b>1.00</b>	<b>1.00</b>	<b>0.00</b>	[0.00, 0.12]	-		
		Biology	<b>Biology (n=263)</b>	2	<b>1.00</b>	<b>1.00</b>	<b>0.01</b>	[0.00, 0.12]	-		
	MLR	Preparatory	Science (n=139)	10	0.84	0.52	0.17	[0.09, 0.26]	<b>0.04</b>		
		Chemistry	<b>Chemistry (n=137)</b>	1	<b>1.00</b>	<b>1.04</b>	<b>0.00</b>	[0.00, 0.09]	<b>0.01</b>		
		General	<b>Science (n=835)</b>	5	<b>0.99</b>	<b>0.98</b>	<b>0.04</b>	[0.00, 0.07]	<b>0.01</b>		
		Chemistry	<b>Chemistry (n=855)</b>	3	<b>1.00</b>	<b>0.99</b>	<b>0.03</b>	[0.00, 0.06]	<b>0.01</b>		
		General	<b>Science (n=258)</b>	2	<b>1.00</b>	<b>1.02</b>	<b>0.00</b>	[0.00, 0.07]	<b>0.02</b>		
		Biology	<b>Biology (n=263)</b>	3	<b>0.99</b>	<b>0.97</b>	<b>0.05</b>	[0.00, 0.10]	<b>0.02</b>		
		Career items ( <i>df</i> = 2)	WLSMV	Preparatory	<b>Science (n=139)</b>	0	<b>1.00</b>	<b>1.00</b>	<b>0.00</b>	[0.00, 0.09]	-
				Chemistry	Chemistry (n=137)	3	<b>1.00</b>	<b>1.00</b>	0.07	[0.00, 0.20]	-
General	Science (n=835)			22	<b>1.00</b>	<b>0.99</b>	0.11	[0.07, 0.15]	-		
Chemistry	Chemistry (n=855)			22	<b>1.00</b>	<b>1.00</b>	0.11	[0.07, 0.15]	-		
General	Science (n=258)			5	<b>1.00</b>	<b>0.99</b>	0.08	[0.07, 0.17]	-		
MLR	Biology		Biology (n=263)	5	<b>1.00</b>	<b>1.00</b>	0.07	[0.00, 0.16]	-		
	Preparatory		<b>Science (n=139)</b>	1	<b>1.00</b>	<b>1.02</b>	<b>0.00</b>	[0.00, 0.07]	<b>0.01</b>		
	Chemistry		<b>Chemistry (n=137)</b>	2	<b>1.00</b>	<b>1.00</b>	<b>0.02</b>	[0.00, 0.15]	<b>0.01</b>		
	General		Science (n=835)	15	<b>0.97</b>	<b>0.92</b>	0.09	[0.06, 0.12]	<b>0.02</b>		
	Chemistry		Chemistry (n=855)	10	<b>0.99</b>	<b>0.98</b>	0.07	[0.04, 0.11]	<b>0.01</b>		
	General	Science (n=258)	11	<b>0.95</b>	0.85	0.14	[0.08, 0.20]	<b>0.03</b>			
	Biology	<b>Biology (n=263)</b>	4	<b>1.00</b>	<b>0.99</b>	<b>0.06</b>	[0.00, 0.13]	<b>0.01</b>			

<sup>a</sup>Acceptable data-model fit values differ by estimator: for WLSMV cut-off values are CFI and TLI  $\geq 0.95$  and RMSEA  $\leq 0.05$ ; for MLR cut-off values are CFI and TLI  $\geq 0.95$ , RMSEA  $\leq 0.06$ , and SRMR  $\leq 0.08$ .



## Reliability

The results of the single-factor models provided in Table 4 were used to determine the course and wording conditions under which each scale showed acceptable data model fit with either the WLSMV or MLR estimator. The models for the intrinsic, grade, and career scales are congeneric because the items were not constrained to be associated with the factor to the same extent. Under congeneric model conditions, omega is the most appropriate single-administration reliability coefficient to report (Komperda, Pentecost, *et al.*, 2018). The self-determination and self-efficacy scales with loadings constrained to be equal are tau-equivalent models where alpha is an acceptable single-administration reliability coefficient. Since alpha and omega are identical under tau-equivalent conditions, only omega values are reported in Table 5.

**Table 5.** Single-administration reliability values (omega) by course, wording, and scale. No value is reported for scales that did not have acceptable single-factor data-model fit.

Course	Preparatory Chemistry		General Chemistry		General Biology	
Wording	Science	Chemistry	Science	Chemistry	Science	Biology
Intrinsic	–	–	–	–	–	–
Self-determination	0.81	–	–	–	0.82	0.86
Self-efficacy	0.89	–	–	0.86	–	–
Grade	–	0.90	0.90	0.88	0.89	0.90
Career	0.92	0.92	–	–	–	0.95

No omega value is reported in Table 5 for scales that did not meet the previously determined data-model fit criteria for each estimator (CFI and TLI  $\geq 0.95$  and RMSEA  $\leq 0.05$  for WLSMV; CFI and TLI  $\geq 0.95$ , RMSEA  $\leq 0.06$ , and SRMR  $\leq 0.08$  for MLR). Though the omegas reported in Table 5 for the modified scales are generally higher than previously reported reliability values for the scales (Glynn *et al.*, 2011; Salta and Koulougliotis, 2015; Schumm and Bogner, 2016; Schmid and Bogner, 2017; Ardura and Pérez-Bitrián, 2018; Komperda, Hosbein, *et al.*, 2018), they should not be interpreted as providing any indication of scale quality on their own, as there is no set threshold for acceptable reliability (Arjoon *et al.*, 2013; Taber, 2018).

Rather, the reliability values are one of many pieces of evidence that should be evaluated to provide overall evidence for the quality of data obtained from an instrument.

### **Limitations**

Though this study was designed as a follow up to previous examination of the functioning of the SMQ II in introductory and general chemistry courses with both the science and chemistry wording (Komperda, Hosbein, *et al.*, 2018), the current study did not obtain enough participants from introductory chemistry courses to test mSMQ II functioning in that population. Future work should obtain data from a more diverse population of students to provide more generalizable information about the functioning of the instrument. Similarly, though previous studies had interviewed biology students to modify items (Glynn *et al.*, 2011) the current study only interviewed general chemistry students to examine patterns in interpretation of the SMQ II items therefore additional interviews with other populations of students would be beneficial.

### **Conclusions**

The goal of this study was to identify potential reasons for the inconsistent internal structure validity evidence for the SMQ II in published literature and determine if modifications to address these issues would improve the internal structure of the instrument. Though the modifications undertaken in this study ultimately resulted in acceptable data-model fit for the overall five-factor model in five out of six wording and course conditions when using an estimator appropriate for the data characteristics, WLSMV, the need for such extensive revisions here and in other studies (Tosun, 2013; Salta and Koulougliotis, 2015; Kwon, 2016; González *et al.*, 2017; Ardura and Pérez-Bitrián, 2018) suggests there may be deeper issues with the underlying theoretical framework of the SMQ II. None of the individual scales demonstrated

acceptable functioning across all course and wording conditions and the best functioning scale had the least theoretical support from the motivation literature. These results further support the interpretation that the items themselves are not well aligned with motivational theories and that calculation of individual scale scores provides little meaningful and interpretable information. It would also not be acceptable to create a total motivation score given the varied theoretical nature of the individual scales.

Though the grade motivation scale had relatively better fit than other scales, this is not an aspect of motivation with a strong theoretical foundation in the self-regulatory literature. The grade scale was created by the developers during revision from the original SMQ to the SMQ II based on EFA results and interviews with students (Glynn *et al.*, 2011; Table A1). The original grade (and career) items were initially intended to belong to a single factor representing extrinsic motivation, a more theoretically-based aspect of motivation (Ryan and Deci, 2000) but evidence from this study does not suggest the grade and career items belong to a single factor. In addition to the theoretical concerns for the grade scale, there are practical concerns in how useful this scale is to instructors and researchers since students were likely to always select responses on the far end of the response scale resulting in a ceiling effect.

In contrast, the self-efficacy scale has more theoretical justification and when tested with constraints, had acceptable data-model fit in some course and wording conditions and more students making use of the entire response scale. The original set of items on the self-efficacy scale were also found to function well when translated to Spanish and used with the physics wording (González *et al.*, 2017). For this reason, the self-efficacy scale may be useful to some practitioners, though it would be beneficial for future research to explore adding additional items to the scale.

It is likely that students value the science content they are learning for multiple reasons and these could potentially have some overlap that needs to be addressed. This would explain the incongruence of the intrinsic and career items overlapping so strongly when intrinsic and extrinsic motivation are on opposite ends of a motivational continuum as described by self-determination theory (Ryan and Deci, 2000). In recent years, Brophy (2008) noted that students' motivation to learn has been examined in three categories: 1) the influence of the classroom milieu, 2) students' expectancies or self-beliefs, and 3) their perceived value of the task. The SMQ II tends to focus more on the latter two and thus the expectancy-value model (Wigfield and Eccles, 2000) might be more harmonious with the intended purpose of the measure. For example, the items for the intrinsic, career, and grade motivation scales on the SMQ II are more aligned with different task value perspectives than the polar ends of the extrinsic-intrinsic continuum. In the expectancy-value model, Eccles and Wigfield (2002) articulate different reasons for why an individual might value a task. For the SMQ II, it tends to be that the two most related might be utility (e.g., the usefulness of a chemistry course for reaching one's career goals) and intrinsic value (because the content is inherently interesting to learn about). When breaking down the intrinsic, career, and grade motivation scales, we could potentially see that a student inherently finds chemistry to be interesting to learn about and thus wants to pursue a career in chemistry. Thus, the student is interested in attaining good grades in her chemistry course based on the relevance to her future career pursuits. For this student, her motivation would be somewhere between 'identified' or 'integrated' regulation on Ryan and Deci's (2000) extrinsic-intrinsic continuum.

## Implications for researchers

When examining academic motivation, it is important to consider its multifaceted nature. What research has shown is that expectancies (e.g., self-efficacy beliefs) are more related to direct performance within a particular course, while having a lower relationship to things like major or career choice in the future (Harackiewicz and Hulleman, 2010). On the other hand, task value is not as highly related to course performance, yet shows greater predictive value of future choice decisions and persistence in a major (e.g., Harackiewicz *et al.*, 2008). Therefore, as educational researchers we need to carefully consider why we are using particular scales and ensure alignment of scales to our research objectives.

With specific regard to the SMQ II and its discipline-specific variants (i.e., BMQ II, CMQ II, PMQ II and others), researchers deciding to use this assessment instrument are encouraged to carefully consider the outcomes of this study, our prior work (Komperda, Hosbein, *et al.*, 2018), and the number of SMQ II studies compiled within this manuscript when analyzing their data. Given the repeated lack of substantiation for the structure of the SMQ II, researchers choosing to use the instrument are urged to continue conducting single and multi-factor CFAs to evaluate the structure of their data.

While psychometric studies can provide insights regarding *if* an assessment instrument, administered in a specific context, produces valid and reliable data, they are not necessarily designed to explain *why* validity or reliability might not be supported. Therefore, when structural validity issues arise, such as those that have been observed with the SMQ II, qualitative studies designed specifically to explore the underlying theoretical framework might be warranted. Consequently, it is recommended that future work on the SMQ II include cognitive interviews using open-ended probes designed to elicit the nature of an underlying construct (e.g., intrinsic

motivation) within a specific context (e.g., chemistry and biology courses). These studies would provide an understanding beyond the item-level cognitive interviews conducted to support response process validity and might provide insight to the salient features of a construct and the way in which cuing of subjects affects student ideas and therefore, why the same structural model might not be equally supported.

### **Implications for practitioners**

As this study and several others by different researchers have been unsuccessful in supporting a key aspect of validity (i.e. structural) for the SMQ II, practitioners are highly cautioned in selecting this tool for use in their evaluation of classroom practices. While the SMQ II and its variations are highly prevalent in the literature, there are other academic motivation instruments and scales available for use in higher education STEM learning environments. These include scales based on motivational theories such as expectancy-value theory (Ferrell and Barbera, 2015; Flake *et al.*, 2015; Kosovich *et al.*, 2015) and self-determination theory (Black and Deci, 2000; Hall and Webb, 2014; Jenó *et al.*, 2017; Liu *et al.*, 2017) and other more general student motivation measures (Pintrich *et al.*, 1991). Overall, it is best to identify an instrument that is most closely aligned with a variable of interest (e.g., self-efficacy) and has evidence for acceptable functioning in environments most similar to those in which it will be used (e.g., discipline, course, target populations). Using assessment instruments that have been shown to produce valid and reliable data in similar environments provides some level of support for the interpretations of the data produced by the instrument. If an assessment instrument is not available, practitioners are encouraged to consider collaborating with an education researcher to study and/or adapt one of the currently available instruments for use in their learning environment. These types of collaborations are symbiotic as the practitioner acquires an

assessment tool aligned with their specific needs and the researcher and larger chemistry education community gains new insights to the available tools being used to understand our educational landscape.

## Appendix

**Table A1.** Trajectory of scales and item modifications from the original SMQ to the SMQ II (Glynn and Koballa, 2006; Glynn *et al.*, 2009, 2011)

2006 (Original SMQ)		2009 (SMQ Modifications)		2011 (SMQ II)	
<b>Intrinsically Motivated Science Learning</b>		<b>Intrinsic Motivation and Personal Relevance</b>		<b>Intrinsic Motivation</b>	I am curious about discoveries in science.
	I enjoy learning the science.		I enjoy learning the science.		Learning science makes my life more meaningful.
	The science I learn is more important to me than the grade I receive.		The science I learn is more important to me than the grade I receive.		I enjoy learning science.
	I find learning the science interesting.		I find learning the science interesting.		Learning science is interesting.
	I like science that challenges me.		I like science that challenges me.		
Understanding the science gives me a sense of accomplishment.	Understanding the science gives me a sense of accomplishment.				
<b>Relevance of Learning Science to Personal Goals</b>	The science I learn relates to my personal goals.	<b>Intrinsic Motivation and Personal Relevance</b>	The science I learn relates to my personal goals.	<b>Intrinsic Motivation</b>	
	I think about how the science I learn will be helpful to me.		I think about how the science I learn will be helpful to me.		
	I think about how I will use the science I learn.		I think about how I will use the science I learn.		
	The science I learn is relevant to my life.		The science I learn is relevant to my life.		The science I learn is relevant to my life.
	The science I learn has practical value for me.		The science I learn has a practical value for me.		
<b>Confidence (Self-efficacy) in Learning Science</b>	I expect to do as well or better than other students in the science course.	<b>Self-efficacy and Assessment Anxiety</b>		<b>Self-efficacy</b>	I am confident I will do well on science labs and projects.
	I am confident I will do well on the science labs and projects.		I am confident I will do well on the science labs and projects.		I believe I can master science knowledge and skills.
	I believe I can master the knowledge and skills in the science course.		I believe I can master the knowledge and skills in the science course.		I am confident I will do well on science tests.
	I am confident I will do well on the science tests.		I am confident I will do well on the science tests.		I believe I can earn a grade of "A" in science.
	I believe I can earn a grade of "A" in the science course.		I believe I can earn a grade of "A" in the science course.		I am sure I can understand science.
<b>Anxiety About Science Assessment</b>	I am nervous about how I will do on the science tests.	<b>Self-efficacy and Assessment Anxiety</b>	I am nervous about how I will do on the science tests.	<b>Self-efficacy</b>	
	I become anxious when it is time to take a science test.		I become anxious when it is time to take a science test.		
	I worry about failing the science tests.		I worry about failing the science tests.		
	I am concerned that the other students are better in science.		I am concerned that the other students are better in science.		
	I hate taking the science tests.		I hate taking the science tests.		



	2006 (Original SMQ)	2009 (SMQ Modifications)	2011 (SMQ II)
<b>Responsibility (Self-determination) for Learning Science</b>	If I am never having trouble learning the science, I try to figure out why.	If I am never having trouble learning the science, I try to figure out why.	
	I put enough effort into learning the science.	I put enough effort into learning the science.	I put enough effort into learning science.
	I use strategies that ensure I learn the science well.	I use strategies that ensure I learn the science well.	I use strategies to learn science well.
	It is my fault if I do not understand the science.		
	I prepare well for the science tests and labs.	I prepare well for the science tests and labs.	I prepare well for science tests and labs.
<b>Extrinsically Motivated Science Learning</b>			I study hard to learn science.
			I spend a lot of time learning science.
	I think about how learning the science can help me get a good job.	I think about how learning the science can help me get a good job.	
	I think about how learning the science can help my career.	I think about how learning the science can help my career.	
			Learning science will help me get a good job.
			Understanding science will benefit me in my career.
			Knowing science will give me a career advantage
			I will use science problem-solving skills in my career.
			My career will involve science.
<b>Grade Motivation</b>	Earning a good science grade is important to me.	Earning a good science grade is important to me.	Getting a good science grade is important to me
	I think about how my science grade will affect my overall grade point average.	I think about how my science grade will affect my overall grade point average.	
	I like to do better than the other students on the science tests.	I like to do better than the other students on the science tests.	I like to do better than other students on science tests.
		I expect to do as well as or better than other students in the science course.	
		It is my fault, if I do not understand the science.	
			I think about the grade I will get in science.
			Scoring high on science tests and labs matters to me.
		It is important that I get an "A" in science.	

**Table A2.** Summary of SMQ II studies utilizing exploratory factor analysis.

Citation	Sample	Modifications	Rotation	Method	Number of Factors
Glynn <i>et al.</i> , 2011	340 university biology students	None	Varimax and Direct Oblimin	Principal Components Analysis and Principal Axis Factoring	Five
Ardura and Pérez-Bitrián, 2018	530 high school students	Translated to Spanish; Physics and chemistry wording	Varimax	Principal Components Analysis	Five
Austin <i>et al.</i> , 2018	2648 university chemistry students	Organic chemistry wording; Removed six items	Promax	Principal Axis Factoring	Four
Kwon, 2016	334 middle school students	Technology wording; Changed response scale; Removed six items	Varimax	Principal Axis Factoring	Five
Schmid and Bogner, 2017	209 high school students	Changed response scale; Only used self-efficacy, self-determination, and career motivation scales; Reduced scales to four items each	Oblimin	Principal Axis Factoring	Three
Schumm and Bogner, 2016	226 high school students	Adapted for German; Removed three items	Oblique	Principal Axis Factoring	Five
Yamamura and Takehira, 2017	165 pharmacy students	Translated to Japanese; Pharmacy wording; Removed 12 items including all self-efficacy items	Promax	Maximum Likelihood	Four

**Table A3.** Summary of SMQ II studies utilizing confirmatory factor analysis with acceptable data-model fit index values bolded.

Citation	Sample	Modifications	*Fit Indices			Model	Estimator
			CFI	RMSEA	SRMR		
Glynn <i>et al.</i> , 2011	340 university biology students	None	0.91	0.07	<b>0.04</b>	Correlated five-factor	Not reported
Ardura & Pérez-Bitrián, 2018	530 high school students	Translated to Spanish; Physics and chemistry wording	0.92	0.07	<b>0.04</b>	Correlated five-factor	Not reported
González, Fernández, & Paoloni, 2017	520 high school students	Translated to Spanish; Physics wording; Only self-efficacy scale	<b>0.998</b>	<b>0.043</b>	<b>0.011</b>	Single factor	Not reported
Komperda, Hosbein, <i>et al.</i> , 2018	146 university general chemistry students	None	0.94	0.08	Not reported	Correlated five-factor	WLSMV
Komperda, Hosbein, <i>et al.</i> , 2018	141 university general chemistry students	Chemistry wording	<b>0.96</b>	0.08	Not reported	Correlated five-factor	WLSMV
Komperda, Hosbein, <i>et al.</i> , 2018	189 university introductory chemistry students	None	<b>0.97</b>	0.07	Not reported	Correlated five-factor	WLSMV
Komperda, Hosbein, <i>et al.</i> , 2018	184 university introductory chemistry students	Chemistry wording	0.94	0.09	Not reported	Correlated five-factor	WLSMV
Kwon, 2016	334 middle school students	Technology wording; Changed response scale	0.939	0.067	Not reported	Correlated five-factor	Not reported
Salta and Koulougliotis, 2015	330 high school students	Translated to Greek; Chemistry wording; Removed lab references	0.91	<b>0.06</b>	<b>0.06</b>	Correlated five-factor	ML

Citation	Sample	Modifications	<sup>a</sup> Fit Indices			Model	Estimator
			CFI	RMSEA	SRMR		
Tosun, 2013	306 high school students	Translated to Turkish; Chemistry wording; Removed six items	<b>0.96</b>	<b>0.059</b>	Not reported	Correlated five-factor	Not reported
Tosun, 2013	266 university students	Translated to Turkish; Chemistry wording; Removed five items	<b>0.96</b>	<b>0.059</b>	Not reported	Correlated five-factor	Not reported
Vasques <i>et al.</i> , 2018	203 university students (Pre)	Translated to Japanese; Removed a self-efficacy item	0.86	0.10	Not reported	Correlated five-factor	Not reported
Vasques <i>et al.</i> , 2018	230 university students (Post)	Translated to Japanese; Removed a self-efficacy item	0.86	0.11	Not reported	Correlated five-factor	Not reported
Yamamura and Takehira, 2017	165 pharmacy students	Translated to Japanese; Pharmacy wording; Removed 12 items including all self-efficacy items	<b>0.994</b>	<b>0.020</b>	Not reported	Correlated four-factor with cross loadings	Not reported

<sup>a</sup>Fit index values recommended by Hu and Bentler (1999) are CFI  $\geq 0.95$ ; RMESA  $\leq 0.06$  and SRMR  $\leq 0.08$ .

**Table A4.** Percentages of experts (n=12) indicating alignment of each SMQ II item with a response scale. Cells with a majority of expert selections are bolded.

Item	SMQ II wording	Frequency (%)	Likert (%)	Either (%)
Intrinsic	I1 The science I learn is relevant to my life	8%	<b>58%</b>	33%
	I2 Learning science is interesting	25%	33%	42%
	I3 Learning science makes my life more meaningful	8%	<b>75%</b>	17%
	I4 I am curious about discoveries in science	25%	33%	42%
	I5 I enjoy learning science	8%	33%	<b>58%</b>
Self-determination	SD1 I put enough effort into learning science	<b>67%</b>	0%	33%
	SD2 I use strategies to learn science well	<b>58%</b>	0%	42%
	SD3 I spend a lot of time learning science	42%	25%	33%
	SD4 I prepare well for science tests and labs	<b>67%</b>	0%	33%
	SD5 I study hard to learn science	<b>67%</b>	0%	33%
Self-efficacy	SE1 I am confident I will do well on science tests	42%	25%	33%
	SE2 I am confident I will do well on science labs and projects	42%	25%	33%
	SE3 I believe I can master science knowledge and skills	25%	<b>50%</b>	25%
	SE4 I believe I can earn a grade of “A” in science	25%	<b>58%</b>	17%
	SE5 I am sure I can understand science	17%	<b>58%</b>	25%
Grade	G1 I like to do better than other students on science tests	17%	<b>58%</b>	25%
	G2 Getting a good science grade is important to me	8%	<b>58%</b>	33%
	G3 It is important that I get an A in science	17%	<b>75%</b>	8%
	G4 I think about the grade I will get in science	<b>75%</b>	8%	17%
	G5 Scoring high on science tests and labs matters to me	25%	42%	33%
Career	C1 Learning science will help me get a good job	0%	<b>92%</b>	8%
	C2 Knowing science will give me a career advantage	0%	<b>92%</b>	8%
	C3 Understanding science will benefit me in my career	0%	<b>83%</b>	17%
	C4 My career will involve science	0%	<b>92%</b>	8%
	C5 I will use science problem-solving skills in my career	17%	42%	42%

**Table A5.** Codes assigned to student responses to SMQ II items. Italicized portions of the examples represent language reflective of the respective code.

Code – Definition	Example quote	Response	Item
<b>Frequency-based</b> Used words from the frequency scale or similar time-based words.	“I feel like I take to chemistry pretty well. <i>Occasionally</i> there's things I struggle with, things I need to study harder with. But I feel like I understand chemistry well and <i>frequently</i> do well on my tests and stuff.”	Usually	SE1
	“I'm pretty confident, you know, in my skills with chemistry. So, it's <i>not always</i> on my mind....”	Sometimes	G4
<b>Quantity-based</b> Used language that involved quantity or comparison in their response that was not specifically time-based.	“I really love science, but chemistry seems to be most math for me. Math is, like, the scariest thing I can imagine. If I imagine hell, it's just doing math over and over again. Yeah, so it's sometimes because <i>I love science but not the math part.</i> ”	Sometimes	I5
	“I think it kinda relates to the same question earlier, where it does give a career advantage because it does open up <i>more opportunities for more open jobs, compared to the average person.</i> So that's why I thought that it would give you the career advantage.”	Usually	C2

**Table A6.** Exemplar student responses to SMQ II items during interviews.

Item	Wording	Student Response
I3	Learning science makes my life more meaningful	“Yeah, I think it provides, it's part of the understanding of what's in front of me, that kind of stuff that I think. I don't know. I just like understanding and I think that's an important part of my life.”
		“Yeah. I feel it really helps out in a lot of ways and especially just an understanding of the world around you and stuff.”
G4	I think about the grade I will get in science	“All the time. Yeah. I mean, I'm still thinking about going into the health field, and they stress grades. And so I have to think about it all the time.”
		“All the time, that's just another thing, I gotta get good grades to get into the program. To get a good job.”
SE2	I am confident I will do well on science labs and projects	“I did pretty well on my labs. We don't have any kind of projects in gen chem”
		“I'm pretty confident with my science labs. I'm not sure about projects because I don't think I've done any.”
		“Yeah. The reason why that wasn't usually is the labs part, I think it's more difficult for me to feel like I will do well on labs because personally, I feel like I don't know what to expect as much as just the tests, but I still think it's possible for me to do well on them.”
		“Usually, because I think the labs compared to the tests are more give and take, and you can interact with our T.A, and see what you did wrong immediately and get feedback. So I think usually I'll do well on those with the occasional difficulty.”

**Table A7.** Descriptive statistics for phase 2 mSMQ II data.

Item	Course	Wording	n	Mean	St. dev.	Median	Min	Max	Skew	Kurtosis
I1	Prep Chem	Science	0	--	--	--	--	--	--	--
		Chemistry	137	3.71	1.02	4	1	5	-0.56	-0.13
	Gen Chem	Science	245	4.18	0.94	4	1	5	-1.09	0.73
		Chemistry	852	3.42	1.19	4	1	5	-0.42	-0.67
	Gen Bio	Science	19	4.58	0.61	5	3	5	-1.17	0.58
Biology		261	3.93	1.12	4	1	5	-0.96	0.25	
I1a	Prep Chem	Science	139	4.35	0.80	5	2	5	-1.13	0.74
		Chemistry	0	--	--	--	--	--	--	--
	Gen Chem	Science	832	4.32	0.87	5	1	5	-1.34	1.53
		Chemistry	261	3.8	1.06	4	1	5	-0.69	-0.28
	Gen Bio	Science	258	4.43	0.81	5	1	5	-1.65	3.22
Biology		20	3.95	1.19	4	1	5	-1.14	0.63	
I2	Prep Chem	Science	139	4.48	0.81	5	1	5	-1.93	4.40
		Chemistry	137	3.89	1.08	4	1	5	-1.07	0.77
	Gen Chem	Science	835	4.43	0.85	5	1	5	-1.72	3.02
		Chemistry	855	3.82	1.12	4	1	5	-0.89	0.09
	Gen Bio	Science	258	4.52	0.80	5	1	5	-2.20	5.86
Biology		263	4.2	1.00	4	1	5	-1.48	2.04	
I3a	Prep Chem	Science	139	4.37	0.92	5	1	5	-1.54	2.05
		Chemistry	137	3.86	1.04	4	1	5	-0.66	-0.09
	Gen Chem	Science	835	4.46	0.80	5	1	5	-1.65	2.95
		Chemistry	855	3.84	1.12	4	1	5	-0.88	0.07
	Gen Bio	Science	258	4.47	0.73	5	2	5	-1.29	1.15
Biology		263	4.18	1.03	4	1	5	-1.40	1.53	
I3b	Prep Chem	Science	139	4.26	0.96	5	1	5	-1.19	0.85
		Chemistry	137	3.54	1.14	4	1	5	-0.40	-0.61
	Gen Chem	Science	835	4.33	0.88	5	1	5	-1.37	1.70
		Chemistry	855	3.65	1.14	4	1	5	-0.61	-0.35
	Gen Bio	Science	258	4.37	0.80	5	1	5	-1.17	0.99
Biology		263	4.03	1.08	4	1	5	-0.94	0.19	
I4	Prep Chem	Science	139	4.39	0.83	5	1	5	-1.53	2.42
		Chemistry	137	3.72	1.11	4	1	5	-0.71	-0.07
	Gen Chem	Science	835	4.41	0.84	5	1	5	-1.62	2.73
		Chemistry	855	3.67	1.16	4	1	5	-0.67	-0.37
	Gen Bio	Science	258	4.39	0.86	5	1	5	-1.47	2.07
Biology		263	3.97	1.07	4	1	5	-1.06	0.62	
I5	Prep Chem	Science	139	4.39	0.88	5	1	5	-1.56	2.38
		Chemistry	137	3.64	1.13	4	1	5	-0.69	-0.19
	Gen Chem	Science	835	4.35	0.88	5	1	5	-1.51	2.20
		Chemistry	855	3.52	1.25	4	1	5	-0.55	-0.70
	Gen Bio	Science	258	4.43	0.85	5	1	5	-1.97	4.63
Biology		263	4.01	1.11	4	1	5	-1.02	0.25	
SD1a	Prep Chem	Science	139	4.5	0.64	5	2	5	-1.27	1.96
		Chemistry	137	4.31	0.82	4	1	5	-1.29	1.84
	Gen Chem	Science	835	4.41	0.73	5	1	5	-1.30	2.11
		Chemistry	855	4.24	0.83	4	1	5	-1.17	1.49
	Gen Bio	Science	258	4.51	0.70	5	1	5	-1.56	3.01
Biology		263	4.4	0.82	5	1	5	-1.53	2.40	

Item	Course	Wording	n	Mean	St. dev.	Median	Min	Max	Skew	Kurtosis
SD2	Prep Chem	Science	139	4.17	0.84	4	2	5	-0.71	-0.30
		Chemistry	137	3.83	0.95	4	1	5	-0.59	-0.03
	Gen Chem	Science	835	4.07	0.9	4	1	5	-0.73	-0.06
		Chemistry	855	3.72	1.00	4	1	5	-0.64	-0.06
	Gen Bio	Science	258	4.09	0.89	4	2	5	-0.78	-0.10
Biology		263	3.98	0.91	4	1	5	-0.78	0.34	
SD3	Prep Chem	Science	139	4.35	0.75	4	1	5	-1.21	2.10
		Chemistry	137	3.74	1.20	4	1	5	-0.76	-0.27
	Gen Chem	Science	835	4.34	0.81	5	1	5	-1.13	0.77
		Chemistry	855	4.07	0.95	4	1	5	-0.90	0.23
	Gen Bio	Science	258	4.33	0.83	5	1	5	-1.33	1.86
Biology		263	4.11	0.97	4	1	5	-1.15	0.87	
SD4a	Prep Chem	Science	139	4.13	0.90	4	1	5	-1.11	1.27
		Chemistry	137	3.99	1.02	4	1	5	-1.09	0.83
	Gen Chem	Science	835	3.91	0.94	4	1	5	-0.81	0.40
		Chemistry	855	3.76	1.01	4	1	5	-0.69	-0.06
	Gen Bio	Science	258	4.07	0.91	4	1	5	-1.12	1.31
Biology		263	4	0.99	4	1	5	-0.94	0.52	
SD5	Prep Chem	Science	139	4.47	0.73	5	2	5	-1.35	1.53
		Chemistry	137	4.13	0.99	4	1	5	-1.19	1.04
	Gen Chem	Science	835	4.31	0.84	5	1	5	-1.27	1.47
		Chemistry	855	4.1	0.94	4	1	5	-0.97	0.46
	Gen Bio	Science	258	4.36	0.83	5	1	5	-1.31	1.40
Biology		263	4.27	0.88	4	1	5	-1.24	1.21	
SD5a	Prep Chem	Science	139	4.47	0.74	5	2	5	-1.33	1.36
		Chemistry	137	4.07	1.00	4	1	5	-1.05	0.70
	Gen Chem	Science	835	4.48	0.77	5	1	5	-1.6	2.68
		Chemistry	855	4.4	0.83	5	1	5	-1.55	2.46
	Gen Bio	Science	258	4.46	0.82	5	1	5	-1.70	2.87
Biology		263	4.4	0.80	5	1	5	-1.51	2.77	
SE1	Prep Chem	Science	139	3.65	1.18	4	1	5	-0.79	-0.21
		Chemistry	137	3.45	1.28	4	1	5	-0.58	-0.75
	Gen Chem	Science	835	3.41	1.13	4	1	5	-0.37	-0.68
		Chemistry	855	3.13	1.24	3	1	5	-0.23	-0.98
	Gen Bio	Science	258	3.5	1.15	4	1	5	-0.57	-0.44
Biology		263	3.24	1.23	3	1	5	-0.32	-0.86	
SE2a	Prep Chem	Science	139	3.92	1.05	4	1	5	-0.83	-0.16
		Chemistry	137	3.83	1.15	4	1	5	-1.09	0.46
	Gen Chem	Science	835	3.77	1.01	4	1	5	-0.64	-0.18
		Chemistry	855	3.63	1.15	4	1	5	-0.62	-0.49
	Gen Bio	Science	258	3.83	1.01	4	1	5	-0.81	0.14
Biology		263	3.66	1.15	4	1	5	-0.68	-0.33	
SE3a	Prep Chem	Science	139	4.12	0.93	4	1	5	-1.07	0.98
		Chemistry	137	3.8	1.21	4	1	5	-0.94	0.01
	Gen Chem	Science	835	3.94	1.04	4	1	5	-0.82	0.02
		Chemistry	855	3.61	1.18	4	1	5	-0.62	-0.52
	Gen Bio	Science	258	3.97	1.01	4	1	5	-0.98	0.47
Biology		263	3.84	1.09	4	1	5	-0.84	0	



Item	Course	Wording	n	Mean	St. dev.	Median	Min	Max	Skew	Kurtosis
SE4a	Prep Chem	Science	139	4	1.12	4	1	5	-1.18	0.73
		Chemistry	137	3.96	1.16	4	1	5	-1.22	0.76
	Gen Chem	Science	835	3.88	1.08	4	1	5	-0.81	-0.11
		Chemistry	855	3.68	1.21	4	1	5	-0.67	-0.54
	Gen Bio	Science	258	4.1	1.01	4	1	5	-1.16	0.79
Biology		263	3.76	1.24	4	1	5	-0.84	-0.34	
SE5	Prep Chem	Science	139	4.25	0.89	4	1	5	-1.39	2.01
		Chemistry	137	3.99	1.04	4	1	5	-1.22	1.12
	Gen Chem	Science	835	4.18	0.89	4	1	5	-1.07	0.92
		Chemistry	855	3.84	1.08	4	1	5	-0.85	0.01
	Gen Bio	Science	258	4.29	0.86	4	1	5	-1.42	2.22
Biology		263	4.06	1.01	4	1	5	-1.12	0.75	
G1a	Prep Chem	Science	139	4.33	0.82	5	2	5	-0.92	-0.17
		Chemistry	137	4.21	1.01	5	1	5	-1.39	1.76
	Gen Chem	Science	835	4.31	0.88	5	1	5	-1.08	0.45
		Chemistry	855	4.16	0.97	4	1	5	-1.02	0.47
	Gen Bio	Science	258	4.22	0.97	5	1	5	-1.19	1.04
Biology		263	4.13	1.01	4	1	5	-1.06	0.60	
G2	Prep Chem	Science	139	4.84	0.45	5	2	5	-3.44	13.92
		Chemistry	137	4.83	0.49	5	1	5	-4.45	27.37
	Gen Chem	Science	835	4.78	0.51	5	1	5	-2.86	11.00
		Chemistry	855	4.76	0.52	5	1	5	-2.50	7.74
	Gen Bio	Science	258	4.86	0.41	5	2	5	-3.33	13.14
Biology		263	4.82	0.51	5	1	5	-4.19	24.41	
G3a	Prep Chem	Science	139	4.79	0.50	5	3	5	-2.41	5.04
		Chemistry	137	4.66	0.70	5	1	5	-2.91	10.82
	Gen Chem	Science	835	4.68	0.62	5	1	5	-2.26	6.07
		Chemistry	855	4.64	0.66	5	1	5	-2.14	5.16
	Gen Bio	Science	258	4.79	0.50	5	2	5	-2.89	10.66
Biology		263	4.74	0.57	5	1	5	-3.11	14.04	
G4	Prep Chem	Science	139	4.79	0.53	5	2	5	-3.11	11.16
		Chemistry	137	4.7	0.66	5	1	5	-3.22	13.73
	Gen Chem	Science	835	4.74	0.58	5	1	5	-2.70	9.04
		Chemistry	855	4.74	0.60	5	1	5	-2.71	8.41
	Gen Bio	Science	258	4.86	0.38	5	3	5	-2.72	7.06
Biology		263	4.73	0.61	5	1	5	-3.15	12.8	
G4a	Prep Chem	Science	139	4.47	0.94	5	1	5	-1.82	2.37
		Chemistry	137	4.2	1.20	5	1	5	-1.48	1.09
	Gen Chem	Science	835	4.51	0.89	5	1	5	-2.12	4.32
		Chemistry	855	4.47	0.90	5	1	5	-1.92	3.34
	Gen Bio	Science	258	4.53	0.86	5	1	5	-2.20	4.79
Biology		263	4.52	0.88	5	1	5	-2.23	4.95	
G5a	Prep Chem	Science	139	4.81	0.43	5	3	5	-2.09	3.70
		Chemistry	137	4.73	0.61	5	1	5	-3.09	12.41
	Gen Chem	Science	835	4.71	0.60	5	1	5	-2.64	9.24
		Chemistry	855	4.71	0.60	5	1	5	-2.56	8.53
	Gen Bio	Science	258	4.81	0.45	5	3	5	-2.29	4.68
Biology		263	4.72	0.63	5	1	5	-3.00	11.42	

Item	Course	Wording	n	Mean	St. dev.	Median	Min	Max	Skew	Kurtosis
C1	Prep Chem	Science	139	4.55	0.76	5	1	5	-1.89	3.88
		Chemistry	137	3.99	1.02	4	1	5	-1.0	0.66
	Gen Chem	Science	835	4.53	0.75	5	1	5	-1.87	3.94
		Chemistry	855	3.73	1.23	4	1	5	-0.69	-0.50
Gen Bio	Science	258	4.52	0.79	5	2	5	-1.63	1.92	
	Biology	263	4.05	1.10	4	1	5	-1.07	0.44	
C2	Prep Chem	Science	139	4.56	0.82	5	1	5	-2.34	5.92
		Chemistry	137	4.04	1.05	4	1	5	-1.10	0.80
	Gen Chem	Science	835	4.56	0.72	5	1	5	-1.83	3.59
		Chemistry	855	3.87	1.15	4	1	5	-0.84	-0.14
Gen Bio	Science	258	4.62	0.73	5	1	5	-2.22	5.51	
	Biology	263	4.17	1.10	5	1	5	-1.39	1.29	
C3	Prep Chem	Science	139	4.59	0.84	5	1	5	-2.44	6.29
		Chemistry	137	4.09	1.08	4	1	5	-1.20	0.96
	Gen Chem	Science	835	4.64	0.69	5	1	5	-2.16	4.96
		Chemistry	855	3.87	1.22	4	1	5	-0.89	-0.20
Gen Bio	Science	258	4.68	0.78	5	1	5	-3.01	9.64	
	Biology	263	4.35	1.00	5	1	5	-1.78	2.69	
C4	Prep Chem	Science	139	4.42	1.02	5	1	5	-1.83	2.61
		Chemistry	137	3.85	1.25	4	1	5	-0.83	-0.35
	Gen Chem	Science	835	4.59	0.78	5	1	5	-2.19	4.90
		Chemistry	855	3.57	1.35	4	1	5	-0.60	-0.84
Gen Bio	Science	258	4.66	0.79	5	1	5	-2.93	9.14	
	Biology	263	4.11	1.25	5	1	5	-1.39	0.84	
C5	Prep Chem	Science	139	4.41	0.91	5	1	5	-1.68	2.54
		Chemistry	137	3.73	1.16	4	1	5	-0.75	-0.20
	Gen Chem	Science	835	4.53	0.76	5	1	5	-1.88	4.07
		Chemistry	855	3.74	1.20	4	1	5	-0.76	-0.31
Gen Bio	Science	258	4.47	0.81	5	1	5	-1.86	4.17	
	Biology	263	4.05	1.14	4	1	5	-1.14	0.47	

## Exploratory factor analysis of phase one data

### *Methods*

Due to the low number of responses from students enrolled in preparatory chemistry courses, only data from general chemistry and general biology courses were used to create the training dataset for the exploratory factor analysis (EFA), leaving the intact preparatory chemistry data as cross validation for testing the confirmatory factor models. The R package *caret* (version 6.0-80; Kuhn, 2008) was used with the general biology and general chemistry data to create two equal partitions of data for each course and wording condition with the restriction

of attempting to create equivalent gender distributions in each partition. The EFA was conducted on these two training datasets using functions available in the psych package (version 1.8.4; Revelle, 2018). Prior to conducting the EFA, the data sets for each wording and course conditions were checked for suitability using the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett’s test of sphericity, as was done in the most recent development of the SMQ II by the original developers (Glynn *et al.*, 2011).

Given the inherent non-normality of data collected on a five-point Likert-type scale, along with the fact that the mSMQ II descriptive statistics showed that all five scale points were not being utilized, especially on the grade scale that showed high skew and kurtosis, the data were analyzed with polychoric correlations when used in EFA. Both principal components and principal axis factoring (PAF) methods were used in prior SMQ II research and the results were reported to be similar (Glynn *et al.*, 2011). Therefore, PAF was used for the mSMQ II data. Decisions about the number of factors to retain were made based on having eigenvalues greater than 1, as in previous SMQ II studies (Glynn *et al.*, 2011), as well as the results of parallel analysis (Bandalos and Finney, 2010) with polychoric correlations. Oblique rotation (oblimin) was selected for this analysis since previous research had demonstrated that the motivation factors were correlated (Glynn *et al.*, 2011).

## ***Results***

Exploratory factor analysis (EFA) was not conducted on the preparatory chemistry data due to low sample size. Instead all of the preparatory chemistry data was reserved for the later confirmatory factor analysis (CFA) allowing it to serve as an independent cross validation dataset. For the general chemistry and biology training datasets, results of KMO and Bartlett’s test were similar to results reported by the original developers (Table A8). These tests indicate

that the data were acceptable for EFA with an overall KMO above 0.70 for all course and wording conditions and highly statistically significant Bartlett's tests (Field *et al.*, 2012). For the general biology data, these tests were run excluding the intrinsic item in which the majority of the item response data were missing due to the survey deployment issue (item I1 for the science wording and I1a for the biology wording). For all EFAs, a five-factor solution was reasonable based on eigenvalues greater than one or the results of parallel analysis.

**Table A8.** Results of tests of mSMQ II data suitability for exploratory factor analysis.

Class	General Chemistry		General Biology	
	Science	Chemistry	Science	Biology
Wording				
n	418	429	130	133
KMO	0.76	0.91	0.84	0.89
Bartlett's	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$

As a result of using oblique rotation, the EFA results provide distinct sets of pattern and structure coefficients, therefore, the term 'loading' is not used to avoid confusion (Henson and Roberts, 2006; Bandalos and Finney, 2010). Since the mSMQ II factors were moderately to strongly correlated (majority falling between 0.25 and 0.60), the structure coefficients were more difficult to interpret than the pattern coefficients, as each factor had a strong relation with all items. Therefore, only the values for the pattern coefficients for each EFA are plotted in Figure A1 and the factor correlations are provided in Tables A9 and A10.



**Figure A1.** Pattern coefficients from exploratory factor analysis (EFA) using principle axis factoring and oblique rotation with mSMQ II data by course and wording conditions. Items are ordered as in Table A1 and asterisks indicate items removed after EFA.

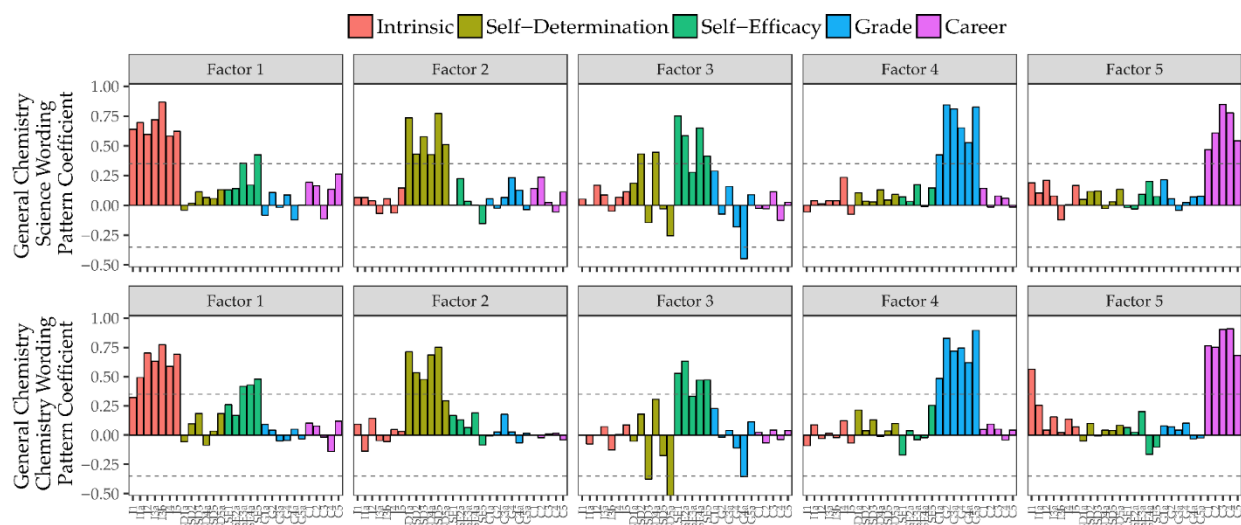
**Table A9.** EFA factor correlations for general chemistry data. Science wording in upper diagonal, chemistry wording in lower diagonal.

	I	SD	SE	G	C
Intrinsic	1	0.44	0.50	0.37	0.66
Self-determination	0.19	1	0.27	0.54	0.44
Self-efficacy	0.35	0.05	1	0.13	0.28
Grade	0.26	0.51	0.02	1	0.44
Career	0.61	0.26	0.15	0.32	1

**Table A10.** EFA factor correlations for general biology data. Science wording in upper diagonal, chemistry wording in lower diagonal.

	I	SD	SE	G	C
Intrinsic	1	0.34	0.45	0.30	0.54
Self-determination	0.39	1	0.19	0.32	0.26
Self-efficacy	0.43	0.23	1	0.26	0.33
Grade	0.37	0.59	0.12	1	0.41
Career	0.61	0.33	0.22	0.35	1

In Figure A1, each of the extracted factors is represented with its own plot showing the pattern coefficients for each item on that factor. Items are ordered along the x-axis based as they appear in Table A1 and also color coded by the scale of their intended association. The developers of the SMQ II used a cutoff of 0.35 to signify that an item was associated with its intended scale to an acceptable degree (Glynn *et al.*, 2011), therefore, a dashed line representing this value (both positive and negative) is shown in Figure A1. For the general biology plots in Figure A1, the intrinsic item with missing data (I1 or I1a) is not plotted depending on the wording condition. To confirm that the low associations between the intrinsic factor and items I1 and I1a were not artifacts of the missing data for these items, separate EFAs were conducted using only a subset of the training data where all students saw both I1 and I1a; only the general chemistry course data provided enough sample size for these calculations (see manuscript Table 2). Results from these EFAs were similar to the larger dataset and are provided in Tables A11 and A12 as well as Figure A2.



**Figure A2.** Exploratory factor analysis results for a subset of general chemistry data in which both I1 and I1a were viewed by participants. Items are ordered as in Table A1.

**Table A11.** Results of tests of data suitability for exploratory factor analysis with subset of general chemistry data responding to both I1 and I1a.

Wording	Science	Chemistry
n	123	133
KMO	0.86	0.83
Bartlett's	$p < 0.001$	$p < 0.001$

**Table A12.** EFA factor correlations for subset of general chemistry data viewing both I1 and I1a. Science wording in upper diagonal, chemistry wording in lower diagonal.

	I	SD	SE	G	C
Intrinsic	1	0.38	0.44	0.36	0.63
Self-determination	0.25	1	0.17	0.51	0.49
Self-efficacy	0.29	0.03	1	0.13	0.22
Grade	0.15	0.38	-0.13	1	0.49
Career	0.43	0.18	-0.07	0.32	1

The EFA results were used to identify potentially problematic items that should be removed before moving into a confirmatory framework. Items were determined to be problematic if they showed low relation to their intended scale factor, if they showed evidence of association with more than one factor, or if they displayed an inconsistent pattern of association with a factor across different wording and course conditions. This last condition is particularly

important for the mSMQ II since the original SMQ II is intended to be used to measure motivation in different contexts. Items identified as problematic are indicated in Table A1 and Figure A1, with an asterisks below their coefficient bar for the factor they were intended to be associated with.

On the intrinsic scale, I1 had a strong association ( $> 0.35$ ) with the career factor in the chemistry wording condition, replicating the concerns that led to its rewording as I1a. Those concerns being that the theoretical framework for the SMQ II identified intrinsic and career motivation as distinct constructs. Additionally, the career motivation scale grew out of a previous extrinsic motivation scale (Glynn *et al.*, 2009) and intrinsic and extrinsic motivation represent opposite ends of the self-determination continuum (Ryan and Deci, 2000). Of additional concern was that the association between I1 and the career factor was not consistent across course and wording conditions. Though neither I1 nor I1a had pattern coefficients below the 0.35 threshold, they were lower in the chemistry courses than in the biology courses providing additional evidence of inconsistent functioning. As a result of these concerns, I1 and I1a were removed.

Three self-determination items were identified for removal due to strong association with the self-efficacy factor. Items SD2, SD4a, and SD5a had the most pronounced association with self-efficacy in the biology courses when seen with the science wording. While SD2 and SD4a had a positive association with the self-efficacy factor above 0.35, item SD5a had a negative association stronger than  $-0.35$ . Since these three items were functioning inconsistently across course and wording conditions, and in some cases were functioning more similarly to self-efficacy items, from which they are intended to be theoretically distinct constructs, they were removed.



The self-efficacy item SE3a showed low association with the self-efficacy factor in all conditions. The association was particularly low for both wordings in the biology courses and the science wording in the chemistry courses. Item SE5 had a similar, though less pronounced, pattern of association and in the general chemistry courses showed an association over 0.35 with the intrinsic scale. These two self-efficacy items were removed.

Two items on the grade scale, G1a and G4a, showed poor association with the grade factor. For both items when seen with the science wording by biology students the association was below the 0.35 threshold. The revised item, G4a, “I worry about my science grade” showed an unintended strong negative association with the self-efficacy factor, an indication that it may be measuring a lack of self-efficacy, which further strengthened its case for removal. On the career scale, item C5 had the lowest association with the career factor and in the biology wording condition had a strong association with the intrinsic factor. Therefore, items G1a, G4a and C5 were removed as well. The items removed as a result of the EFA are indicated in Table A1, leaving 19 items to be tested through confirmatory factor analysis.

## References

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, (2014), *Standards for educational and psychological testing*, Washington, DC: American Educational Research Association.
- Ardura D. and Pérez-Bitrián A., (2018), The effect of motivation on the choice of chemistry in secondary schools: adaptation and validation of the Science Motivation Questionnaire II to Spanish students. *Chem. Educ. Res. Pract.*, **19**(3), 905–918.
- Arjoon J. A., Xu X., and Lewis J. E., (2013), Understanding the state of the art for measurement in chemistry education research: Examining the psychometric evidence. *J. Chem. Educ.*, **90**(5), 536–545.
- Austin A. C., Hammond N. B., Barrows N., Gould D. L., and Gould I. R., (2018), Relating motivation and student outcomes in general organic chemistry. *Chem. Educ. Res. Pract.*, **19**(1), 331–341.
- Bandalos D. L. and Finney S. J., (2010), Factor analysis: exploratory and confirmatory., in Hancock G. R. and Mueller R. O. (eds.), *The reviewer's guide to quantitative methods in the social sciences*. New York: Routledge, pp. 93–114.
- Bandura A., (1993), Perceived self-efficacy in cognitive development and functioning. *Educ. Psychol.*, **28**(2), 117–148.
- Barbera J. and VandenPlas J. R., (2011), All assessment materials are not created equal: the myths about instrument development, validity, and reliability., in Bunce D. M. (ed.), *Investigating classroom myths through research on teaching and learning*. Washington, DC: ACS Symposium Series; American Chemical Society, pp. 177–193.
- Beatty P., (2004), The dynamics of cognitive interviewing., in Presser S., Rothgeb J. M., Couper M. P., Lessler J. T., Martin E., Martin J., and Singer E. (eds.), *Methods for testing and evaluating survey questionnaires*. Hoboken, NJ: Wiley-Interscience, pp. 45–66.
- Beauducel A. and Herzberg P. Y., (2006), On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Struct. Equ. Model. A Multidiscip. J.*, **13**(2), 186–203.
- Black A. E. and Deci E. L., (2000), The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Sci. Educ.*, **84**(6), 740–756.
- Brophy J., (2008), Developing students' appreciation for what is taught in school. *Educ. Psychol.*, **43**(3), 132–141.
- Cagande J. L. L. and Jugar R. R., (2018), The flipped classroom and college physics students' motivation and understanding of kinematics graphs. *Issues Educ. Res.*, **28**(2), 288–307.
- Campos-Sánchez A., López-Núñez J. A., Carriel V., Martín-Piedra M.-Á., Sola T., and Alaminos M., (2014), Motivational component profiles in university students learning histology: a comparative study between genders and different health science curricula. *BMC Med. Educ.*, **14**, 46.
- Chan J. Y. K. and Bauer C. F., (2014), Identifying at-risk students in general chemistry via cluster analysis of affective characteristics. *J. Chem. Educ.*, **91**(9), 1417–1425.
- Childers G. and Jones M. G., (2017), Learning from a distance: High school students' perceptions of virtual presence, motivation, and science identity during a remote microscopy investigation. *Int. J. Sci. Educ.*, **39**(3), 257–273.
- Cleveland L. M., Olimpo J. T., and DeChenne-Peters S. E., (2017), Investigating the relationship

- between instructors' use of active-learning strategies and students' conceptual understanding and affective changes in introductory biology: A comparison of two active-learning environments. *CBE Life Sci. Educ.*, **16**(2), 1–10.
- Cronbach L. J., (1951), Coefficient alpha and the internal structure of tests. *Psychometrika*, **16**(3), 297–334.
- Deci E. L. and Ryan R. M., (2000), The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychol. Inq.*, **11**(4), 227–268.
- Eccles J. S. and Wigfield A., (2002), Motivational beliefs, values, and goals. *Annu. Rev. Psychol.* **2002.**, **53**, 109–132.
- Ferrell B. and Barbera J., (2015), Analysis of students' self-efficacy, interest, and effort beliefs in general chemistry. *Chem. Educ. Res. Pract.*, **318**(16), 318–337.
- Ferrell B., Phillips M. M., and Barbera J., (2016), Connecting achievement motivation to performance in general chemistry. *Chem. Educ. Res. Pract.*, **17**(4), 1054–1066.
- Field A., Miles J., and Field Z., (2012), *Discovering statistics using R*, Los Angeles: SAGE.
- Finney S. J. and DiStefano C., (2013), Non-normal and categorical data in structural equation modeling., in Hancock G. R. and Mueller R. O. (eds.), *Structural equation modeling: a second course*. Charlotte, NC: Information Age Publishing, pp. 439–492.
- Flake J. K., Barron K. E., Hulleman C., McCoach B. D., and Welsh M. E., (2015), Measuring cost: The forgotten component of expectancy-value theory. *Contemp. Educ. Psychol.*, **41**, 232–244.
- Fortus D., (2014), Attending to affect. *J. Res. Sci. Teach.*, **51**(7), 821–835.
- Gadermann A. M., Guhn M., and Zumbo B. D., (2012), Estimating ordinal reliability for likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Pract. Assessment, Res. Eval.*, **17**(3), 1–13.
- Glynn S. M., Brickman P., Armstrong N., and Taasoobshirazi G., (2011), Science motivation questionnaire II: Validation with science majors and nonscience majors. *J. Res. Sci. Teach.*, **48**(10), 1159–1176.
- Glynn S. M. and Koballa T. R., (2006), Motivation to learn in college science., in Mintzes J. J. and Leonard W. H. (eds.), *Handbook of college science teaching*. Arlington, VA: National Science Teachers Association, pp. 25–32.
- Glynn S. M., Taasoobshirazi G., and Brickman P., (2007), Nonscience majors learning science: A theoretical model of motivation. *J. Res. Sci. Teach.*, **44**(8), 1088–1107.
- Glynn S. M., Taasoobshirazi G., and Brickman P., (2009), Science motivation questionnaire: Construct validation with nonscience majors. *J. Res. Sci. Teach.*, **46**(2), 127–146.
- González A., Fernández M.-V. C., and Paoloni P.-V., (2017), Hope and anxiety in physics class: Exploring their motivational antecedents and influence on metacognition and performance. *J. Res. Sci. Teach.*, **54**(5), 558–585.
- González A. and Paoloni P.-V., (2015), Perceived autonomy-support, expectancy, value, metacognitive strategies and performance in chemistry: a structural equation model in undergraduates. *Chem. Educ. Res. Pract.*, **16**(3), 640–653.
- Grosslight L., Unger C., Jay E., and Smith C. L., (1991), Understanding models and their use in science: Conceptions of middle and high school students and experts. *J. Res. Sci. Teach.*, **28**(9), 799–822.
- Hall N. and Webb D., (2014), Instructors' support of student autonomy in an introductory physics course. *Phys. Rev. Spec. Top. - Phys. Educ. Res.*, **10**(2), 1–22.

- Hallgren K. A., (2012), Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor. Quant. Methods Psychol.*, **8**(1), 23–34.
- Harackiewicz J. M., Durik A. M., Barron K. E., Linnenbrink-Garcia L., and Tauer J. M., (2008), The role of achievement goals in the development of interest: Reciprocal relations between achievement goals, interest, and performance. *J. Educ. Psychol.*, **100**(1), 105–122.
- Harackiewicz J. M. and Hulleman C. S., (2010), The importance of interest: The role of achievement goals and task values in promoting the development of interest. *Soc. Personal. Psychol. Compass*, **4**(1), 42–52.
- Henson R. K. and Roberts J. K., (2006), Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educ. Psychol. Meas.*, **66**(3), 393–416.
- Hibbard L., Sung S., and Wells B., (2016), Examining the effectiveness of a semi-self-paced flipped learning format in a college general chemistry sequence. *J. Chem. Educ.*, **93**(1), 24–30.
- Hu L. and Bentler P. M., (1999), Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model. A Multidiscip. J.*, **6**(1), 1–55.
- Jeno L. M., Raaheim A., Kristensen S. M., Kristensen K. D., Hole T. N., Haugland M. J., and Mæland S., (2017), The relative effect of team-based learning on motivation and learning: A self-determination theory perspective. *CBE Life Sci. Educ.*, **16**(4), 1–12.
- Kassae A. M., (2016), Examining the role of motivation and mindset in the performance of college students majoring in STEM fields. Middle Tennessee State University.
- Komperda R., Hosbein K. N., and Barbera J., (2018), Evaluation of the influence of wording changes and course type on motivation instrument functioning in chemistry. *Chem. Educ. Res. Pract.*, **19**(1), 184–198.
- Komperda R., Pentecost T. C., and Barbera J., (2018), Moving beyond alpha: A primer on alternative sources of single-administration reliability evidence for quantitative chemistry education research. *J. Chem. Educ.*, **95**(9), 1477–1491.
- Kosovich J. J., Hulleman C. S., Barron K. E., and Getty S., (2015), A practical measure of student motivation: Establishing validity evidence for the expectancy-value-cost scale in middle school. *J. Early Adolesc.*, **35**(5–6), 790–816.
- Krosnick J. A. and Presser S., (2010), Question and questionnaire design., in Marsden P. V. and Wright J. D. (eds.), *Handbook of survey research*. Bingley, UK: Emerald Group Publishing Limited.
- Kuhn M., (2008), Building predictive models in R using the caret package. *J. Stat. Softw.*, **28**(5), 1–26.
- Kwon H., (2016), Effect of middle school students' motivation to learn technology on their attitudes toward engineering. *Eurasia J. Math. Sci. Technol. Educ.*, **12**(9), 2281–2294.
- Liu Y., Ferrell B., Barbera J., and Lewis J. E., (2017), Development and evaluation of a chemistry-specific version of the academic motivation scale (AMS-Chemistry). *Chem. Educ. Res. Pract.*, **18**(1), 191–213.
- Liu Y., Raker J. R., and Lewis J. E., (2018), Evaluating student motivation in organic chemistry courses: moving from a lecture-based to a flipped approach with peer-led team learning. *Chem. Educ. Res. Pract.*, **19**, 251–264.
- Mahrou A., Ginger K., and Rowell H., (2016), Motivationally-informed interventions for at-risk STEM students. *J. STEM Educ.*, **17**(3), 77–84.

- National Research Council, (2012), *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*, Singer S. R., Nielsen N. R., and Schweingruber H. A. (eds.) Washington, DC: The National Academies Press.
- Olimpo J. T., Fisher G. R., and Dechenne-Peters S. E., (2016), Development and evaluation of the tigrionus course-based undergraduate research experience: Impacts on students' content knowledge, attitudes, and motivation in a majors introductory biology course. *CBE Life Sci. Educ.*, **15**(4), 1–15.
- Peters G.-J. Y., (2017), userfriendlyscience: Quantitative analysis made accessible, [Computer Software].
- Pintrich P. R., Smith D. A. F., Garcia T., and McKeachie W. J., (1991), *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*, Ann Arbor, MI.
- R Core Team, (2019), R: A language and environment for statistical computing, [Computer Software].
- Reece A. J. and Butler M. B., (2017), Virtually the same: A Comparison of STEM students content knowledge, course performance, and motivation to learn in virtual and face-to-face introductory biology laboratories. *J. Coll. Sci. Teach.*, **46**(3), 83–89.
- Revelle W., (2018), psych: Procedures for psychological, psychometric, and personality research, [Computer Software].
- Riccitelli M., (2015), Science identity's influence on community college students engagement, persistence, and performance in biology. Northcentral University.
- Rosseel Y., (2012), lavaan: An R package for structural equation modeling. *J. Stat. Softw.*, **48**(2), 1–36.
- Ryan R. M. and Deci E. L., (2000), Self-determination theory and the facilitation of intrinsic motivation. *Am. Psychol.*, **55**(1), 68–78.
- Salta K. and Koulougliotis D., (2015), Assessing motivation to learn chemistry: Adaptation and validation of Science Motivation Questionnaire II with Greek secondary school students. *Chem. Educ. Res. Pract.*, **16**(2), 237–250.
- Schmid S. and Bogner F. X., (2017), How an inquiry-based classroom lesson intervenes in science efficacy, career-orientation and self-determination. *Int. J. Sci. Educ.*, **39**(17), 2342–2360.
- Schumm M. F. and Bogner F. X., (2016), Measuring adolescent science motivation. *Int. J. Sci. Educ.*, **38**(3), 434–449.
- Shin S., Lee J. K., and Ha M., (2017), Influence of career motivation on science learning in Korean high-school students. *Eurasia J. Math. Sci. Technol. Educ.*, **13**(5), 1517–1538.
- Simpkins S. D., Davis-Kean P. E., and Eccles J. S., (2006), Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Dev. Psychol.*, **42**(1), 70–83.
- Srisawasdi N., (2015), Evaluation of motivational impact of a computer-based nanotechnology inquiry learning module on the gender gap. *J. Nano Educ.*, **7**(1), 28–37.
- Taber K. S., (2018), The use of cronbach's alpha when developing and reporting research instruments in science education. *Res. Sci. Educ.*, **48**(6), 1273–1296.
- Tosun C., (2013), Adaptation of Chemistry Motivation Questionnaire-II to Turkish. *J. Educ. Fac.*, **15**(1), 173–202.

- Vasques D. T., Yoshida L., Ellinger J., and Maninang J. S., (2018), Validity and reliability of the science motivation questionnaire II (SMQ II) in the context of a Japanese university., in *New perspectives in science education*. Florence: Pixel.
- Wigfield A. and Eccles J. S., (2000), Expectancy–value theory of achievement motivation. *Contemp. Educ. Psychol.*, **25**(1), 68–81.
- Xia Y. and Yang Y., (2018), RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behav. Res. Methods*, (1999), 1–20.
- Yamamura S. and Takehira R., (2017), Effect of practical training on the learning motivation profile of Japanese pharmacy students using structural equation modeling. *J. Educ. Eval. Health Prof.*, **14**, 2.
- Young A. M., Wendel P. J., Esson J. M., and Plank K. M., (2018), Motivational decline and recovery in higher education STEM courses. *Int. J. Sci. Educ.*, **40**(9), 1016–1033.
- Yu C.-Y., (2002), Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. University of California, Los Angeles.
- Zeyer A., (2018), Gender, complexity, and science for all: Systemizing and its impact on motivation to learn science for different science subjects. *J. Res. Sci. Teach.*, **55**(2), 147–171.
- Zimmerman B. J., (2000), Attaining self-regulation: A social cognitive perspective., in Boekaerts M., Pintrich P. R., and Zeidner M. (eds.), *Handbook of self-regulation*. San Diego, CA: Academic Press, pp. 13–39.
- Zusho A., Pintrich P. R., and Coppola B., (2003), Skill and will: The role of motivation and cognition in the learning of college chemistry. *Int. J. Sci. Educ.*, **25**(9), 1081–1094.