Investigating Bifactor Modeling of Biology Undergraduates' Task Values and Achievement Goals Across Semesters

Date resubmitted: February 11, 2023

Date Published: August 1, 2023

Greene, J. A., Bernacki, M. L., Plumley, R. D., Kuhlmann, S. L., Hogan, K. A., Evans, M., Gates, K. M., & Panter, A. T. (2023). Investigating bifactor modeling of biology undergraduates' task values and achievement goals across semesters. *Journal of Educational Psychology, 115*(6), 836-858. https://doi.org/10.1037/edu0000803

Abstract

Undergraduate STEM students' motivations have a strong influence on whether and how they will persist through challenging coursework and into STEM careers. Proper conceptualization and measurement of motivation constructs, such as students' expectancies and perceptions of value and cost (i.e., expectancy value theory; EVT) and their goals (i.e., achievement goal theory; AGT), are necessary to understand and enhance STEM persistence and success. Research findings suggest the importance of exploring multiple measurement models for motivation constructs, including traditional confirmatory factor analysis, exploratory structural equation models (ESEM), and bifactor models, but more research is needed to determine whether the same model fits best across time and context. As such, we measured undergraduate biology students' EVT and AGT motivations and investigated which measurement model best fit the data, and whether measurement invariance held across three semesters. Having determined the best-fitting measurement model and type of invariance, we used scores from the best performing model to predict biology achievement. Measurement results indicated a bifactor-ESEM model had the best data-model fit for EVT and an ESEM model had the best data-model fit for AGT, with evidence of measurement invariance across semesters. Motivation factors, in particular attainment value and subjective task value, predicted small to medium-sized amounts of variance in biology course outcomes each semester. Our findings provide support for using modern measurement models to capture students' STEM motivations and potentially refine conceptualizations of them. Such future research will enhance educators' ability to benevolently monitor and support students' motivation, and enhance STEM performance and career success.

Keywords: achievement goal theory; situated expectancy value theory; exploratory structural equation modeling; bifactor modeling; achievement motivation

Educational Impact and Implications Statement

Students' motivations are key predictors of persistence and success in postsecondary STEM courses and majors, yet recent research has suggested a need to use more advanced and complex models for measuring these motivations. Our study extended this previous work by testing these more advanced and complex models in a new course context and across multiple semesters. Our findings supported using these more complex but likely more accurate measurement models to understand postsecondary students' motivations in an introductory biology course. These more accurate measurement models proved helpful for predicting final exam scores and course grades, with implications for how researchers conceptualize the many aspects of student motivation. These findings highlight the importance of carefully theorizing and measuring student motivation, because accurate data are needed to help educators create and monitor classrooms that promote positive student motivation. Such measures can also help researchers develop better tools for intervening upon students' motivations, fostering successful STEM career trajectories and success.

Investigating Bifactor Modeling of Biology Undergraduates' Task Values and Achievement Goals Across Semesters

There is increasing demand for STEM professionals in the workforce (National Science Board, 2021), yet post-secondary educators have struggled to increase STEM graduation rates (Olson & Riordan, 2012), in part due to significant student attrition out of STEM majors (Chen, 2015). Such attrition likely has many causes, ranging from the cultural (e.g., racialized experiences in STEM; Eccles & Wigfield, 2020), to the contextual (e.g., unsupportive classroom climates; Seymour et al., 2019), to the individual (e.g., feelings of disenchantment with STEM; Rosenzweig et al., 2021), all likely dynamically interacting with one another. Such interactions often manifest in decreasing STEM motivations over time (Cromley et al., 2016; Perez et al., 2019; Robinson et al., 2019), prompting significant interest in ways of conceptualizing, measuring, and monitoring these motivations over the course of students' STEM career (Jiang et al., 2020). Such monitoring is necessary to understand how to construct cultures, contexts, and individual supports that promote STEM persistence, achievement, and careers. Yet, understanding STEM motivation has proven challenging due to the many theories and models of motivation in the literature (Wigfield & Koenka, 2020).

Conceptualizations of motivation range from a focus on students' expectancies and perceptions of value and cost (i.e., expectancy value theory; EVT; Eccles & Wigfield, 2020), to their goals (i.e., achievement goal theory; AGT; Elliot & Dweck, 2005), to students' sense of autonomy and agency or lack thereof (i.e., self-determination theory; Ryan & Deci, 2000), among others. These theories have driven a large corpus of empirical research into understanding and promoting positive STEM course and career motivations (Cromley et al., 2016; Hong & Bernacki, 2022; Lee et al., 2022; Perez et al., 2019; Robinson et al., 2019, 2022; Rosenzweig et al., 2020). Likewise, there has been a great deal of recent scholarship on how best to model responses to motivation instruments, ranging from more traditional confirmatory factor analysis methods (e.g., Muenks et al., 2023) to recently resurgent bifactor methods (Holzinger & Swineford, 1937; Lohbeck et al., 2022) and profile analyses (Lee et al., 2022). Recent STEM education research supporting the utility of using bifactor-exploratory structural equation modeling (ESEM) measurement models with EVT conceptualizations (Lohbeck et al., 2022; Part et al., 2020, 2023) suggests the need to consider refining EVT. However, before embarking on such theory development, it is important test whether such measurement models and their theoretical implications are reproduced in new contexts, with different samples, and across multiple time points (Greene, 2022). In addition, evidence supporting the use of bifactor-ESEM measurement models with EVT indicates a need to test similar measurement models with other motivation conceptualizations (i.e., achievement goal theory). Thus, more research is needed to determine the optimal measurement model for motivation constructs, across contexts and time, which, in turn, can be used to refine theory and derive scores that provide more accurate and useful information about students' motivation (Robinson et al., 2022).

Therefore, in this study, we gathered college students' responses to items from multiple motivation conceptualizations (i.e., EVT, AGT), across three consecutive semesters of the same introductory biology course, to determine which measurement model proved most informative for each conceptualization, across each semester. Then, we examined whether and how our derived motivation scores predicted course performance. Our findings extend previous research on modeling motivation to new contexts (i.e., a biology course at a different institution than the one studied in Part et al., 2020), new constructs (i.e., testing bifactor and ESEM models with achievement goal theory), and across course semesters, with implications for how scholars conceptualize and measure EVT and AGT motivations and for how scholars and practitioners use those conceptualizations and measures for monitoring and intervening upon motivation across time and context.

Theories of Achievement Motivation

Theories that describe students' motivation for academic tasks are numerous and consider individuals' beliefs about themselves (e.g., self-concept, self-efficacy), their interests in engaging specific tasks, the value and enjoyment a task could provide, the costs it might require, and how task engagement might suit one's goals, afford or challenge one's autonomy and feelings of relatedness, or arouse desirable or undesirable emotions during learning, among others (Wentzel & Miele, 2018). With an aim to understand the ways learners engage in active learning in hybrid STEM courses, we focused primarily on two theories. First, we examined students' perceptions of themselves as learners poised to engage in tasks that promise value but also can be costly, for which students hold task-specific expectancies (Eccles & Wigfield, 2020). Second, we focused on the achievement goals that compel students' course engagement (Elliot & Dweck, 2005).

Situated Expectancy-Value Theory

Eccles and Wigfield (Eccles et al., 1983; Eccles & Wigfield, 2020) have argued that students' subjective expectancies for success and their perceptions of the value or cost of engaging in tasks determine their motivation, which in turn affects their decisions to engage, complete, and thrive in tasks, including academic activities. The expectancy aspect of situated expectancy-value theory bears some resemblance to self-efficacy (i.e., students' perception of their likely success at completing specific tasks; Bandura, 2001) and is often conceptualized (Wigfield & Eccles, 2000) and measured as such (e.g., Nagengast et al., 2011). Students also have perceptions of the value of specific tasks, which affects their motivation. Value perceptions have four aspects (Eccles & Wigfield, 1983). Attainment value concerns how students view the importance of successfully completing a task, with such value related to their sense of identity. Tasks perceived as strongly related to a student's identity have more attainment value than those seen as more tangential. Intrinsic value concerns how enjoyable a task is to a student. Utility value concerns students' perception of the alignment between a task and their current and future goals. Finally, cost concerns students' perceptions of the negative consequences of engaging in a task. Modern conceptualizations of EVT have included three kinds of cost, based on the amount of effort the task is perceived to take, the opportunities to pursue other tasks that would be lost when engaging in a task, and the negative psychological consequences of engaging in a task (e.g., Perez et al., 2014). Importantly, the various aspects of EVT are subjective (i.e., a task that may have significant attainment value for one student may be perceived as having low attainment value for another). Eccles and Wigfield (2020) have recently asserted an elaboration of their prior work, called situated expectancy value theory, which prioritizes students' context, culture, and history when developing interventions and when determining whether and how findings might generalize from one context, group, or time period to another.

Achievement Goal Theory

Achievement goal theorists have posited students' learning behaviors are initiated and sustained by the goals they set to gain competency (Elliot & Dweck, 2005). According to the perspective of achievement goal theory, students set competency goals that involve a *definition* (Elliot & MacGregor, 2001; Elliot & Murayama, 2008) to either master content or perform well with content (e.g., mastery or performance). These definitions interact with the goal's *valence*, which denotes a desire to approach success or avoid failure (Harackiewicz et al., 2002). As such, achievement goal theorists measure students' endorsement of mastery approach, mastery avoidance, performance approach, and performance avoidance goals. However, theory and research have established mastery avoidance goals tend not to be relevant to tasks that introduce novel content. Instead, these goals are only relevant to instances where one may possess sufficient competency to set goals that avoid declines in such mastery (Van Yperen et al., 2009). Because the early undergraduate STEM course in this study introduces students to central topics in biology for which they are assumed to have little to no prior knowledge, we focused on the measurement of mastery approach, performance approach, and performance avoidance goals.

Students with higher endorsements of mastery approach goals aim to develop competence, which suggests these students are more motivated to actively construct meaning from learning materials. Indeed, findings from prior research suggest mastery approach goals result in more adaptive outcomes, like greater self-efficacy, increased persistence, and effective use of self-regulated learning strategies (Kaplan & Maehr, 2007; Urdan & Kaplan, 2020). On the other hand, students who endorse performance goals seek to demonstrate a desired level of competence in comparison to their peers, which is less likely to motivate them toward actively constructing meaning from learning materials. Typically, performance goals are posited to be less adaptive than mastery goals (Pintrich, 2000), but the approach and avoidance distinctions highlight important nuance. Students with higher endorsements of performance approach goals are oriented toward demonstrating ability, such as performing better than their peers, whereas students with higher endorsements of performance avoidance goals are oriented towards simply not demonstrating poor performance. Thus, performance approach goals can be adaptive for learning, especially when students' performance goals are oriented towards maximizing achievement, and not toward social norms (Dompnier et al., 2013). In contrast, performance avoidance goals are typically less adaptive for learning. For example, researchers have found that performance approach goals are related to increased engagement, interest, and learning (e.g., Senko & Harackiewicz, 2005; Krou et al., 2021; Yeh et al., 2019), whereas performance avoidance goals are associated with lower intrinsic motivation, engagement, and learning (e.g., Church et al., 2001; Pajares et al., 2000; Pekrun et al., 2009). It is important to note, however, that performance approach goals have also been linked to maladaptive outcomes, such as cheating and avoiding help-seeking (e.g., Karabenick, 2004; Tas & Tekkaya, 2010). Finally, students often pursue multiple goals simultaneously (Barron & Harackiewicz, 2001), where concurrently held goals comprise a goal complex which predicts outcomes in different ways than do independently held achievement goals (e.g., Sommet & Elliot, 2017). Taken together, mastery approach, performance approach, and performance avoidance goals are all important constructs when considering how a student will be motivated to pursue competency goals, which in turn affect a range of outcomes, including achievement (Simon et al., 2015).

Modeling Motivation Data

A robust literature has evolved around the question of how best to conceptualize and model situated expectancy value theory (Eccles & Wigfield, 2020). The expectancy aspect of the model is relatively uncontroversial; typically, it is conceptualized as a single latent construct. Likewise, most EVT scholars have modeled value as comprised of three specific constructs or types: attainment, intrinsic, and utility. There is greater variance in how scholars have conceptualized and modeled the cost aspects of EVT (Flake et al., 2015; Perez et al., 2014). Cost has been modeled as a single construct (Eccles et al., 1983) but also as composed of three (i.e., effort, opportunity, and psychological; Perez et al., 2014) or four (i.e., effort, outside effort, loss of valued alternatives, and emotional; Flake et al., 2015) distinct types, and either as independent correlated factors or as indicators of a general cost latent factor (Part et al., 2020).

Differences in views regarding the generality or specificity of cost suggest the need for rigorous investigation of multiple conceptualizations of EVT (Muenks et al., 2023; Part et al., 2020). Such differences may be due, in part, to the statistical modeling approaches used to investigate the construct validity of competing conceptual models of EVT. Typical methods of amassing construct validity evidence are often overly conservative, such as typical confirmatory factor analysis (CFA; see Figure 1) models of item response data where each survey item is allowed to load only on its posited latent factor (e.g., an item designed to measure participant's perceptions of psychological cost is modeled to load only on a psychological cost latent factor and no other factors) and only latent factors are allowed to correlate (Asparouhov, Muthén, & Morin, 2015). These conservative models preclude even small, incidental cross-loadings between items and non-posited factors (e.g., the item designed to indicate psychological cost allowed to cross-load on effort and opportunity cost factors). The omission of small, incidental cross-loadings can lead to biased estimates of factor correlations and increased multicollinearity among latent factor scores (Howard et al., 2018). An alternative to confirmatory factor analysis, exploratory structural equation modeling (ESEM; see Figure 1), removes this restriction and allows each measured indicator to load on each latent factor, both posited and non-posited. Simulation studies suggest ESEM factor correlations tend to be closer to population values than those from CFA models (Asparouhov, Muthén, & Morin, 2015), resulting in factor scores with better discrimination and likely stronger predictive validity with outcomes of interest (e.g., achievement).

However, both CFA and ESEM still require an a priori decision about the level of specificity of the posited factors. Rather than choosing either a general or a specific conceptualization of constructs in a model like EVT, researchers have begun examining bifactor versions of CFA and ESEM, where both a general factor and specific factors are modeled simultaneously (see Figure 1; Chen et al., 2012; Howard et al., 2018; Lohbeck et al. 2022; Part et al., 2020, 2023; Reise, 2012). These models partition item variance into a general factor (e.g., overall subjective task value) and several specific factors whose variance contributes to predicting outcomes of interest above and beyond the general one (e.g., specific forms of utility and cost; Chen et al., 2012). Bifactor models, either within a CFA or ESEM framework, can be used to investigate questions regarding the proper conceptualization and modeling of cost within an EVT framework (i.e., as a single factor or multiple factors) as well as how best to conceptualize the relations among cost and value aspects of that model. Part and colleagues (2020) did just that, testing numerous CFA and ESEM models, with and without general bifactors, finding psychometric evidence supporting a bifactor model with items indicating both a single, general subjective task value bifactor as well as six specific factors: attainment value, intrinsic value, utility value, effort cost, opportunity cost, and psychological cost. Support for a bifactor-ESEM model has implications for how motivation scholars conceptualize EVT, but such conceptualizations would benefit from further investigation with other samples in other contexts, thus prompting our study.

In addition, there is value in testing whether data indicating achievement goals (Harackiewicz et al., 2002) should also be modeled using CFA or ESEM methods, with or without a general bifactor (see Figure 1). Such testing of this motivation theory would logically follow from similar work by Part et al (2020), previously described, as well as similar bifactor investigations conducted with other motivation theories, such as self-determination theory (Gunnell & Gaudreau, 2015). Much like Eccles and Wigfield (1983) originally posited the existence of both general and specific perceptions of a task's value, achievement goal theorists (e.g., Elliot & Dweck, 2005) have characterized all achievement goals as reflecting a general desire to gain competence, and that specific goals are distinguished by their definition and valence (Elliot & McGregor, 2001). The empirical observation of emergent goal complexes and high correlations between achievement goals that share definitions and contrast in their valences (e.g., Elliot & Murayama, 2008) further suggests an underlying generality that bears investigation.

Implications of Bifactor Modeling for Motivation Theory

The recent increase in research on bifactor models of motivation constructs (e.g., Howard et al., 2018; Lohbeck et al., 2022; Part et al., 2020, 2023) is not a revolution but rather a renaissance, with the first bifactor models articulated eighty years ago (Holzinger & Swineford, 1937; Reise, 2012). These models have some analytic advantages over the more common hierarchical models (e.g., Muenks et al., 2023), including the ability to more clearly and simultaneously differentiate the relations between an outcome and both general and specific factors (Chen et al., 2006). However, empirical findings supporting a bifactor model necessarily imply revisiting motivation theory, via a process of epistemic iteration (Greene, 2022). Such reconceptualization has already begun for EVT based on the findings of Part and colleagues (2020, 2023) among others. Eccles and Wigfield (2020) directly referenced Part and colleagues' findings when reflecting upon changes to EVT. They seemingly endorsed further investigation of a general subjective task value (STV) and specific value and cost factors, within a situated perspective:

> …we have not specified exactly how the various components would aggregate to form either the STV of individual achievement-related task or the relative STVs across several different task or activity options available to an individual at one time or over time.

> Instead, we assume the weighting of each of these major components likely varies across developmental time and situations. (Eccles & Wigfield, 2020, p. 6)

If we were to find additional evidence that a bifactor measurement model best fit the EVT data, across multiple samples and semesters, this would further support the reconceptualization of EVT already pursued by Part and colleagues and Eccles and Wigfield.

To our knowledge, bifactor measurement models of AGT have not been investigated. However, researchers have conducted other kinds of analyses suggesting the possibility of both a general goal pursuit factor as well as specific achievement goal factors. Soon after the conceptualization of achievement goal theory, researchers observed strong correlations among achievement goals and investigated the phenomenon where learners simultaneously pursued multiple goals as a goal complex (Barron & Harackiewicz, 2001). This simultaneous pursuit of multiple goals suggests an overall "general" level of goal pursuit as well as more specific goal pursuits that can vary in intensity from one another and the general level, all comprising a "multiple goals perspective." More recently, both the general level of goal pursuit across multiple goals and differences in ipsative shapes of the specific goals pursued by respondents can be seen in additional research on achievement goals that involve profile analyses. For example, several groups of researchers have submitted achievement goal factor scores (i.e., summed or factor-derived scores for participants' mastery approach, performance approach, and performance avoidance goals) to latent profile analyses, often finding a relatively small number of trends in those scores (Holden et al., 2021; Schwinger & Wilds, 2012; Tuominen et al., 2020; Wang et al., 2016). Such consistent trends in the individual AGT factors suggest the possibility of a general AGT factor as well as specific AGT factors, mirroring a bifactor approach. Indeed, such a conceptualization has been incorporated into theory by Elliot (2005), via the term "goal complexes." In a recent review of achievement goal theory conceptual and empirical research, Urdan and Kaplan (2020) argued: "Goal complexes involve the psychological interplay between broad motives and specific objectives, which, together, frame and guide the student's cognitive-affective interpretation of the situation and selection of engagement strategies" (p. 7). This discussion of "broad motives" and "specific objectives" aligns well with a bifactor measurement model's general and specific factors. As such, evidence supporting a bifactor measurement model would imply a reconceptualization of AGT similar to the one currently being pursued in EVT research (e.g., Part et al., 2020, 2023).

Given the plausibility of bifactor-congruent conceptualizations of both EVT and AGT, in this study's first research question, we investigated which of the four models in Figure 1 best fit the data from our EVT instrument, as well as which of those same models best captured responses to our AGT instrument. The concurrent effort to estimate a factor structure for values, costs, and achievement goals extends a line of research where complex models of motivation, from multiple theories, emerged and predicted achievement (Conley, 2012). In addition, we investigated configural (i.e., same factor structure), metric (i.e., equal factor loadings), and scalar (i.e., equal factor loadings and intercepts) measurement invariance models across semesters (Kline, 2015). Configural invariance indicates the foundational design of the instruments (e.g., which items indicate particular latent factors) holds across semesters. Metric invariance is needed to support cross-semester comparisons of relations among motivation latent factors and other constructs (e.g., SRL processing) and outcomes (e.g., course grade), whereas scalar invariance is needed to examine mean differences in those factors across semesters. Determining the optimal data-model fit for each motivation conceptualization increases the likelihood of, and power for, finding relations among motivations and other aspects of our study (Kline, 2015), including achievement outcomes but also various types of learning interventions (e.g., Bernacki et al., 2020), which was another aspect of the larger scholarship within which this study was conducted.

Accounting for Potential Confounds in the Data

Our collection and analysis of motivation data from students enrolled in introductory biology courses was part of a broader initiative to support these students' success using predictive modeling to deliver targeted self-regulated learning (SRL; Greene, 2018) interventions (Bernacki et al., 2020). A series of meta-analyses have provided ample evidence of the benefits that self-regulated learning interventions can confer to those who complete them in face-to-face workshops, classroom trainings, and on digital platforms (e.g., Hattie, 2009; Theobald, 2021). More recent syntheses have documented the promise of providing support for self-regulated learning practices (Zheng, 2016), as well as evidence of the effects of training in such practices (Broadbent & Poon, 2015; Theobald, 2021). A class of digital, self-regulated learning interventions has been developed to target self-regulatory skills that are particularly useful to students who pursue learning objectives involving declarative and conceptual knowledge acquisition as demanded in STEM coursework (Bernacki et al., 2020, 2021; Cogliano et al., 2020, 2022). These brief, digital interventions can be delivered directly within the digital platforms where hybrid STEM courses house learning materials and activities, and they have been found to improve adoption of desirable learning behaviors (Bernacki et al., 2020) and exam and course performance for learners (Bernacki et al., 2020, 2021). However, learners vary in their responsiveness to interventions, which draws into question the role of students' motivations in STEM persistence and achievement, which is the focus of this particular study. To maintain this focus, our predictive validity analyses will account for any SRL intervention effects, but only as controls so that we can best understand the predictive validity of students' motivation, which is the main focus of the second research question for this study.

The Present Study

Students have many different types of motivation, and each can play a role in STEM persistence and achievement (Urdan & Kaplan, 2020). Researchers and educators would benefit from an investigation into (1) the optimal measurement modeling of both EVT and AGT constructs, (2) whether that modeling is invariant across multiple course semesters, and (3) the ways those motivation constructs predict STEM achievement outcomes. Therefore, across three semesters, we also gathered achievement goal theory and expectancy value theory data in the first few weeks of the course. Given past findings regarding best practices for modeling motivation (Hamilton et al., 2018; Part et al., 2020), we tested CFA and ESEM models, with and without general bifactors, thus reproducing scholarship on expectancy value theory (Part et al., 2020) and extending this scholarship to another motivation model, achievement goal theory (Senko et al., 2011). Our first research question involved determining which of the four models shown in Figure 1 best fit the EVT data and the AGT data. We were interested in whether, in every semester, the same measurement model would be the best fitting one. Also, we investigated the measurement invariance of the best-fitting model across semesters (Kline, 2015). For our second research question, we asked whether two course achievement outcomes, a cumulative final exam and the final course grade, related to motivation factor scores derived from the best fitting models determined for our first research question, controlling for any effects from our SRL intervention.

Method

Transparency and Openness

Methods of sample selection, data exclusion, and outcome measures are detailed below. Data, analysis code, and survey instruments are provided at https://osf.io/cdafk/?view_only=4465cf95217147859ea17d45ce9370ad. Analysis code includes models for all Fall 2019 analyses, which can be used to replicate parallel analyses for Fall 2020 and Fall 2021. Data were analyzed using Mplus version 8.8 (Muthén & Muthén, 2017). This study was not preregistered.

Sample and Context

Participants at a large university in the southeastern US were recruited from an introductory biology course with two sections in Fall 2019 (N = 488), three sections in Fall 2020 (N = 432), and two sections in Fall 2021 (N = 549), all utilizing the same syllabus (i.e., learning objectives, assigned textbook and materials, assignments, exam coverage and timing, instructional design) and pedagogy. This was the first required course for biology majors, who typically comprised around 25 percent of the course population each semester. The remainder of the course population included students from a variety of majors, many of which required that the student pass the course with a grade of B- or better. Of note, Fall 2019 and Fall 2021 semester courses were conducted face-to-face, whereas the Fall 2020 semester course was conducted entirely online due to the COVID-19 pandemic. This study was approved by the university's institutional review board.

Procedure

During the second week of the semester, participants completed online instruments with items derived from achievement goal theory (i.e., a trichotomous conceptualization with mastery approach, performance approach, and performance avoidance items) and expectancy value theory (i.e., self-efficacy, attainment value, intrinsic value, utility value, effort cost, opportunity cost, psychological cost). In the Fall 2019 and Fall 2021 semesters, we collected trace data from participants' interactions with the technology platforms and tools used in the course (e.g., learning management system, in-class response system) for the purpose of identifying participants most likely to benefit from our SRL intervention. We submitted trace data from the first few weeks of the course to a previously developed prediction model to differentiate students likely to receive a grade of C+ or worse (i.e., flagged students) from those likely to receive a grade of B- or better (i.e., non-flagged students), because for many of the students in the course, if they received a C+ or worse they would be required to repeat the course before proceeding in their major, a high-stakes detrimental outcome. Flagged participants were randomly assigned to a treatment or control condition. Then, several weeks before the first course exam, flagged treatment participants received one of two versions of a digital "Science of Learning to Learn" intervention designed to improve their learning skills and performance (Bernacki et al., 2020), whereas flagged control participants received a parallel intervention on an alternate assignment focused on the biology course topics.

In Fall 2020, no prediction model was used and instead all participants were randomly assigned to either a treatment or a control condition. Therefore, no participants were "flagged" because no prediction model was used. Of note, in Fall 2021 there were multiple versions of the treatment intervention, but differences among these versions are not the focus of the current study. For the purposes of this study, we treated assigned condition as a covariate to account for these research design features but did not investigate them further.

In each semester, participants completed four course exams, with the first three administered after distinct units of content, and the fourth being a cumulative "final" exam. Final course grades were comprised of final exam scores (25%), two highest unit exams scores (50%), homework (9%), quizzes (9%), and participation (7%). Course instructors received no indication of which of their students consented to participate in the study, nor were they made aware of participants' motivation scores, whether participants were flagged or not, or participants' assigned condition.

Measures

We measured students' value to engage in the task across three subscales (e.g., attainment value, interest value, utility value) and students' perceived cost of engaging in the task across three other subscales (e.g., effort cost, opportunity cost, psychological cost; Perez et al., 2014). Each subscale was comprised of four items that required a response on a 6-point Likert-type scale. Example items included, "Is the amount of effort it will take to do well in this course worthwhile to you?" (attainment value), "Learning the material covered in this course is enjoyable" (interest value), "What I learn in this course will be useful for me later in life" (utility value), "Considering what I want to do with my life, this course is just not worth the effort" (effort cost), "I'm concerned this course may cost me some treasured friendships" (opportunity cost), and "My self-esteem would suffer if I tried in this course and was unsuccessful" (psychological cost).

The Achievement Goals Questionnaire-Revised (Elliot & Murayama, 2008) was used to measure students' goal orientations and includes four subscales. In this study, we used three task-relevant subscales: mastery approach, performance approach, and performance avoidance. Each subscale was comprised of three items that required a response on a 7-point Likert-type scale. Mastery approach items included, "My aim it to completely master the material presented in this class", "My goal is to learn as much as possible", and "I am striving to understand the content in this course as thoroughly as possible." Performance approach items included, "I am striving to do well in comparison to other students", "My aim is to perform well relative to others", and "My goal is to perform better than the other students." Performance avoidance items included, "My goal is to avoid performing poorly compared to others", "I am striving to avoid performing worse than other", and "My aim is to avoid doing worse than other students." Finally, exam and course grade data were imported from university records.

Measurement Models

Similar to other recent publications on motivation (e.g., Hamilton et al., 2018; Lohbeck et al., 2022; Part et al., 2020), we investigated four different ways of modeling the constructs. As shown in Figure 1, which uses achievement goal theory as the exemplar, the traditional CFA model involves each item loading only on its specific factor with no paths to the other two factors, and with correlations among the three factors. The ESEM model allows each item to load on every factor, albeit with the estimation beginning with a target of loading on the specific factor for which it was designed, and all factors to correlate (Morin et al. 2013). The bifactor versions of each model introduce an omnibus general factor (e.g., general goal striving) on which every item is allowed to load while also precluding correlations among the specific factors (i.e., mastery approach, performance approach, performance avoidance), given the assumption that the general factor captures these relations (Reise, 2012). The bifactor-CFA model retains the restriction that items, beyond loading on the general factor, can only load onto the factor for which they were designed. The bifactor-ESEM model mirrors the bifactor-CFA model except that it allows all items to load on all specific factors, again with each item targeted to load most strongly on the specific factor for which it was designed.

We investigated each of the four measurement models for each motivation theory separately. Further, for expectancy value theory, we investigated the models using only the value and cost factors of the theory (Perez et al., 2014). Self-efficacy was modeled using a traditional CFA model.

Missing Data Management

There were three kinds of missing data in our samples. First, our consent process included two options: participants could consent to our use of their motivation survey data and academic outcome data only versus also consenting to participating in our randomized controlled trial (RCT). Participants who did not consent to participating in our RCT could be included in analyses for research question one but could not be included in analyses for research question two because we included treatment condition as a predictor in that analysis. Second, whereas all students who completed the motivation surveys answered all items, some students did not complete any items on the motivation surveys, meaning they could not be included in any of our analyses. Third, each semester, between 2.5% and 4.2% of enrolled students withdrew from the class at some point before the final exam, meaning they did not have a final exam or course grade score. Thus, our analytic sample differed from the total course population and the sample sizes differed between research questions one and two (see Table 1). Given these types of attrition, we investigated the nature of missing data due to not consenting to partake in our RCT, not completing the motivation surveys, and withdrawing from the course.

First, we conducted Little's (1998) MCAR test for all motivation survey item measured variables and our final exam and course grade variables. The chi-square value was statistically significant, rejecting the null hypothesis that our missing data were missing completely at random. Thus, we investigated the nature of the missing data (Enders, 2022). First, if participants completed the motivation survey, there were complete data for all items; the survey did not allow participants to skip a response. Most participants without motivation data simply did not complete the survey, with a small percentage having withdrawn from the course prior to the survey being administered. As shown in Table 1, for our first research question, 16, 14, and 9.4 percent of the motivation survey data were missing across the three Fall semesters. For our second research question, most of the missing data came from participants who did not consent to participate in our RCT but did complete the course, meaning these participants had final exam and course grade data. A smaller percentage of participants did consent to participate in the RCT but withdrew prior to the end of the course, meaning they were missing final exam and course grade data only. For research question two, missing data could be due to: 1) having no motivation survey scores, 2) consent status (i.e., whether or not they consented to participating in our RCT), or 3) withdrawing from the course. Next, we explored relations among these factors.

Using binary logistic regression where withdrawal was coded as one and enrollment throughout the semester coded as zero, we found that participants who did not complete the motivation surveys were more likely to withdraw from the course in Fall 2020 (Odds Ratio = 9.100, $p < .001$) and Fall 2021 (Odds Ratio = 17.143, $p < .001$) but not in Fall 2019 (Odds Ratio = 1.036, $p = .964$). Second, participants who did not complete the motivation surveys had lower final exam and course grades across all semesters (all *p*s < .022, Cohen's *d* ranged from .340 to 1.136) except for final exam grade in Fall 2020, where no statistically detectable difference emerged. Third, consent status did not statistically significantly relate to motivation factor scores except in three cases. In Fall 2019, psychological cost, attainment value, and performance approach scores were negatively related to participants' decision against participating in the RCT. Fourth, consent status did not statistically significantly predict withdrawals, final exam score, or course grade except for final exam grade in Fall 2019, where participants who did not consent to participate in our RCT outperformed those who did consent ($p < .015$, Cohen's $d = .328$).

Given these findings, for Research Question 1, we employed missing data-handling procedures assuming a conditionally missing at random process where missing data were presumed unrelated to missingness after controlling for one or more of our three sources of missing data (i.e., motivation survey completion, consent status, withdrawal). We conducted analyses using Mplus 8.8's FIML estimator with robust standard errors. We included withdrawal status, final exam, and course grades as auxiliary missing data covariates, using Mplus 8.8's "auxiliary" command. However, auxiliary variables are not allowed in the estimation of ESEM models. For analyses for Research Question 2, there were no missing data on our outcome variables, but our analytic sample sizes ranged from 68% to 75% of the total course population due to students' decision to not participate in our RCT, an inability for us to make a prediction about their performance because they did not complete the pretest, or their decision to withdraw from the course (see Table 1).

Data Analysis

All measurement modeling and regression analyses were conducted using Mplus 8.8. Many of our motivation survey measured variables had skewness-to-standard-error and/or kurtosis-to-standard-error ratios with an absolute value of about two, therefore we conducted all analyses with robust maximum likelihood estimation. We made several analytic decisions regarding our measurement modeling and invariance testing for Research Question 1. For our ESEM and Bifactor-ESEM analyses, we used the rotation methods that were the default in Mplus 8.8: geomin and bi-geomin, respectively. The four types of motivation factor models (i.e., CFA, Bifactor-CFA, ESEM, Bifactor-ESEM; see Figure 1) were compared using each configural model's AIC, BIC, SABIC, RMSEA, CFI, and SRMR criteria, looking for convergent evidence to decide upon the best model fit. We looked to choose models with lower AIC, BIC, SABIC, RMSEA, and SRMR values, and higher CFI values (Kline, 2015). After having chosen the best-fitting type of model for each set of motivation constructs (i.e., EVT, AGT, self-efficacy), we examined measurement invariance of that model across semesters. To do this, we examined the chi-square difference test values across configural, metric, and scalar invariance models within types (Kline, 2015). A statistically significant chi-square difference test between the less parsimonious (e.g., configural model) and the more parsimonious (e.g., metric model) was interpreted as evidence to reject the more parsimonious model in favor of the less parsimonious one.

For Research Question 2, our primary focus was whether scores from our motivation measurement models predicted our two primary achievement outcomes: final exam grade and course grade. Examining these relations required accounting for the various conditions to which participants were assigned across the three semesters. Therefore, we included condition assignment as a "control" so that we could assess our motivation factors' unique predictive validity. In our Results section, we do not focus on the findings regarding our various conditions, given they are ancillary to our focus in this study. To conduct these analyses, we saved factor scores from our motivation measurement models and then used those scores as measured variables in regressions of final exam and course grade on those scores and condition status, again using Mplus 8.8.

Results

Descriptive Statistics

Demographic data for our sample appears in Table 2. Across semesters, approximately 25% of the participants were biology majors, between 20 and 25% of the participants identified as first-generation college students, and between 22 and 30% of students identified as under-represented minorities in STEM fields. Exam and course grade averages were consistent across semesters (see Table 3). Motivation survey item response averages were also relatively consistent across semesters and generally adhered to theory. Effort and opportunity cost means were lower on the response scale than psychological cost. Value item means were higher than cost item means. On average, self-efficacy responses were also relatively high on the response scale. Participants more strongly endorsed mastery approach items, compared to performance approach or avoidance items.

Research Question 1: Determining Best Model for Motivation Data

We estimated each of the four models in Figure 1 for the cost and value items from our expectancy value theory instrument, and then our self-efficacy items, and then once more for the achievement goal theory items.

Expectancy Value Theory Modeling

Across the four types of models, the Bifactor-ESEM models best fit our EVT item data, across all metrics (see Table 4). Based on the chi-square difference tests of the three types of measurement invariance across semesters, metric but not scalar invariance was supported. The other data-model fit indices were similar for the metric and scalar models, thus we chose to adopt the Bifactor-ESEM metric measurement invariance model as our final model for EVT. Of note, typically Bifactor-ESEM models are rotated orthogonally, meaning that latent factors do not correlate with one another. Such rotation is not possible with metric or scalar measurement invariance models, except in the first group (i.e., in our study, Fall 2019; Morin et al., 2020). We did investigate strict measurement invariance (i.e., residual covariances fixed across semesters) as well as invariance of latent factor correlations (i.e., all correlations equal to zero) but those models were statistically significantly worse fits to our data than the metric and scalar models.

The factor loadings largely adhered to EVT theory (see Table 5). In each semester, items designed to measure their latent factor loaded strongest on that factor, albeit often with several statistically significant but practically small cross-loadings to other factors. Attainment value was the latent factor with the fewest strong, posited factor loadings. The subjective task value bifactor had strong, mostly statistically significant loadings, except for two of the psychological cost items that were statistically non-significant. Notably, the model estimation process converged on a solution where cost items loaded positively on the subjective task value factor, and value items loaded negatively, meaning that scores on this factor technically should be interpreted as the lack of subjective task value. This is an artifact of the estimation process and does not change the substantive interpretation of the latent factor itself. Latent factor correlations were rotated to zero for the Fall 2019 group and in the Fall 2020 and Fall 2021 groups, all but three correlations were statistically non-significant and close to zero (see Table 6). Latent factor reliabilities ranged from .652 to .885, indicating strong reliability, on average.

Achievement Goal Theory Modeling

For our AGT items, we were able to test only three of the four posited measurement models, because the Bifactor-ESEM model would not successfully converge (see Table 4). Data-model fit indices revealed the ESEM model was the best measurement model. Chi-square difference tests indicated scalar measurement invariance across semesters for the ESEM model, therefore we selected this model as the best for the AGT items. Examination of the factor loadings revealed patterns as expected given how the items were designed (see Table 5). For each semester, and each latent factor, the items with the strongest positive loadings were those designed to measure the latent factor. There were, however, statistically significant but practically small cross-loadings for each factor. Latent factor correlations were typically strong but not sufficient to warrant concern about collinearity (see Table 7). Correlations between mastery approach and performance avoidance latent factors were all small, reaching statistical significance in two but not all three of the semesters. Latent factor reliabilities were strong, ranging from .815 to 1.000.

Self-Efficacy Modeling

Given our five self-efficacy items were designed to all indicate a single latent factor, CFA was the only measurement model possible. As can be seen in Table 4, data-model fit was only adequate by typical standards (e.g., Hu & Bentler, 1999). The addition of a single residual covariance improved data-model fit to common criteria and scalar measurement invariance was supported via the chi-square difference testing. As such, we accepted this measurement model, with scalar invariance, as our final model. Factor loadings were all positive and relatively strong (see Table 5). The latent factor reliability (i.e., Coefficient H) was .79.

Supplemental Analysis of Measurement Invariance Across Major

In response to a reviewer question, we explored whether motivation model measurement invariance was supported between those participants who were Biology majors and those participants who were not. We reran all our analyses with major status as the groups and found the same measurement model as the one determined in our by-semester analyses was the best fit for each motivation theory (i.e., Bifactor-ESEM for EVT; ESEM for AGT; CFA with one residual covariance for Self-Efficacy). Further, we found scalar measurement invariance across Biology major status for our EVT model and AGT model, and metric invariance for our Self-

Efficacy model. In sum, we interpreted these findings as supporting sufficient measurement invariance across Biology major status to use latent factors scores in our analyses for Research Question 2.

Research Question 2: Predicting Final Exam and Course Grade Achievement

To examine whether and how our motivation factors predicted final exam and course grade, we saved factor scores from each of the final measurement models determined via Research Question 1 analyses. We examined the correlations among our motivation factor scores, final exam, and course grade (see Tables 8-10). There were some consistent positive, statistically significant correlations between motivation factors across semesters. For example, psychological cost positively related to performance approach and performance avoidance factors across all three semesters, with relatively similar estimates except for the correlation between psychological costs and performance approach scores, which was lower in Fall 2019 than in the other two semesters. Psychological costs correlated positively with mastery approach factors in only the Fall 2020 semester. Subjective task value was positively related to mastery approach and performance approach factors across all three semesters, with relatively consistent values. Subjective task value was also positively correlated with self-efficacy across all three semesters, with similar correlation values. Self-efficacy was positively and similarly related to mastery approach scores across all three semesters, also. In two of the three semesters, mastery approach scores were positively related to opportunity cost and intrinsic value scores, also. Effort cost and mastery approach scores were statistically, significantly, negatively related in two of the three semesters. In the Fall 2019 semester, there were two statistically significant correlations that were not present in the other semesters (i.e., mastery approach and performance avoidance scores; attainment value and performance avoidance scores), in Fall 2020 there were eight unique correlations, and in Fall 2021 there were four unique correlations. Finally, correlations among motivation latent factors and final exam scores and course grades were relatively small, with none above the absolute value of .202 and statistical significance for only five out of 22 correlations in Fall 2019, two out of 22 for Fall 2020, and nine out of 22 for Fall 2021.

Next, we used measured variable path analysis to regress final exam (see Table 11) and course grade (see Table 12) variables on each set of factor scores and included dummy-coded variables to control for the RCTs we ran in each semester but that were not the focus of this analysis. Further, on the advice of a reviewer, we used multiple groups analysis to examine whether regression coefficients differed across students who were and were not biology majors. In Fall 2019, a path model with all path coefficients constrained to be equal had a statistically non-significant chi-square value, supporting using this more parsimonious model as our final model, compared to a model where path coefficients could vary. In this model, for both biology major and non-majors, subjective task value negatively predicted final exam score, which, given the valence of this factor, should be interpreted as the more subjective task value participants felt toward the course, the higher their predicted outcome score. Final course grade was positively predicted by participants' attainment value score, but negatively related to their utility value score.

The Fall 2020 path model, with all path coefficients constrained to be equal across biology major status, had a statistically significant chi-square value, implying that one or more path coefficients statistically significantly varied across status groups. A series of Wald tests were performed to identify path coefficients that should be allowed to vary across biology major status groups, resulting in five that, when freed, led to statistically significant improvements in model chi-square. A model with only those five path coefficients free to vary across biology major status had a statistically non-significant model chi-square, indicating good data-model fit. For Non-Biology Majors attainment value positively predicted final exam score, whereas for Biology Majors, only utility value predicted final exam score, with a negative relationship. In terms of final course grade, subjective task value negatively predicted final exam score, which, given the valence of this factor, should be interpreted as the more subjective task value participants felt toward the course, the higher their predicted outcome score. The same biology major status differences existed for course grade as they did for final exam grade, and in addition, self-efficacy was differently weighted for each group. However, none of these path coefficients were statistically significant (i.e., there was no detectable relationship between these motivation factors and final course grade), despite each coefficient being statistically significantly different across biology major status.

Finally, for the Fall 2021 cohort, a model with all path coefficients constrained to be equal across biology major status groups had a statistically non-significant chi-square, indicating good data-model fit and precluding an investigation of whether particular path coefficients differed across biology major status. Therefore, we used the fully constrained model as our final model. Opportunity cost and attainment value were positively, statistically significantly related to final exam score. Subjective task value negatively predicted both final exam score and final course grade, which, given the valence of this factor, should be interpreted as the more subjective task value participants felt toward the course, the higher their predicted outcome score.

In terms of the predictive validity of the models, R^2 values in Fall 2019 and Fall 2021 were large (i.e., R^2 ranged from .132 to .213; see Tables 11 and 12), whereas in Fall 2020, the semester when students participated in class online due to COVID-19, Non-Biology Major R^2s were medium (i.e., .060, .054) and Biology Major R^2s were large (i.e., .143, .199). However, those total R^2s represent the combined predictive validity among all predictors, including statistically non-significant motivation latent factor regression coefficients as well as variance predicted by the treatment effects. Therefore, we calculated the unique predictive validity for just the statistically significant latent motivation factors, as well (see Tables 11 and 12). R^2 equivalents for just these statistically significant latent motivation factors ranged from .018 to .047, often characterized as relatively small relations (Gignac & Szodorai, 2016).

Discussion

Given the United States' need to diversify and support the many pathways to STEM careers (Cannady et al., 2014), more research is needed to determine how to optimize STEM education experiences (Rosenzweig et al., 2021). Such optimization is aided by accurate conceptualization, measurement, and monitoring of student motivation, so that researchers and educators can understand what innovations (e.g., active learning instructional designs, self-regulated learning interventions; Bernacki et al., 2020; Eddy & Hogan, 2014) positively affect students' achievement goals, expectancies, values, self-efficacy, and other relevant phenomena. The findings for our first research question revealed that both expectancy value theory and achievement goal theory were better modeled using ESEM methods, compared to traditional CFA methods, aligning with evidence in the motivation methodological literature (e.g., Asparouhov et al., 2015; Hamilton et al., 2018). Likewise, our study extends Part and colleagues' (2020) findings, showing support for bifactor-ESEM modeling for expectancy value theory in a second educational context (i.e., undergraduate biology courses v. undergraduate anatomy and physiology course, each at different institutions), across three separate semesters. Taken together, this study and the Part et al. (2020) study impel the need for continued exploration of the implications of bifactor-ESEM modeling for expectancy value theory, particularly in terms of what the general subjective task value represents, the strength and roles of the general and specific factors within particular contexts, and the degree to which each kind of factor changes as a result of new experiences.

However, before the implications of measurement findings can be used to refine theory, psychometric explanations for unexpected item functioning must be explored (Greene, 2022). We examined the measurement of SEVT using a widely adopted item set that has been developed by theory authors (Battle & Wigfield, 2003), adapted by additional scholars (Perez et al. 2014), and rigorously investigated here and in other scholarship (Part et al. 2020; Muenks et al., 2023). Overall, items designed to measure SEVT specific factors loaded strongly on their target factors and weakly on other factors. Nonetheless, because task value constructs are understood to be situated within a learning task and environment, additional research is warranted to examine how features of academic tasks and the learners who engage in them may influence the way constructs are measured across studies. In particular, our findings suggest the items measuring attainment value may benefit from review, given two of them had relatively low target loadings, with one item also cross-loading onto the intrinsic value factor with standardized values around .340 (see Table 5). Part et al (2020) found similar psychometric concerns with the attainment value items. Further, the construct reliabilities for the effort cost factor were somewhat lower than those for other factors (i.e., .652-.683), but not far below typical metrics for adequate measurement. Mathematically, the range of targeted loadings for this factor had the lowest ceiling (i.e., .619) among all the factors, which likely contributed significantly to the lower reliability for this factor, as well.

The relatively weak relationships between items and the effort cost factor, compared to the stronger loadings between items and the other SEVT factors, may indicate differences in respondents' experiences of effort versus other costs and values. One avenue for future research is to use mixed methodologies (Meyer & Schutz, 2020) to explore what referent respondents use when responding to effort cost items, compared to other items. For example, estimations of psychological cost refer back to features of the task and estimations of opportunity costs refer to respondents' outside commitments. In both these cases, the referents are relatively clear and tangible. Effort cost estimations, on other hand, require respondents to refer to an estimation of how much effortful engagement tasks might require. This is a less obvious estimation, potentially involving interactions between task features and the respondents' individual characteristics. Such effort cost judgments may vary across learners, and perhaps also as a result of additional phenomena such as their prior preparation for the task, as theorized in the SEVT model, as well as respondents' own personal heuristics for making such judgments. Our data did not include measures of these phenomena, therefore in future research they might need to be collected and investigated to understand the relative strength of the effort cost factor loadings.

Finally, our findings regarding the psychometric qualities of the subjective task value factor cohered with those from Part et al. (2020), including the finding that psychological cost did not seem to indicate as strongly on the general factor as other aspects of SEVT. Again, this common finding, across contexts and samples, suggests further need to investigate the design of the psychological cost items. If psychometric investigations of the items fail to identify profitable revisions to the wordings, then it may be the case that conceptualizations of subjective task value must account for lower prevalence of psychological cost in what Part et al. (2020) characterized as accumulation of social and educational experiences into an "internal working valuation model for an individual's STEM courses" (p. 12).

Our data supported using an ESEM measurement model for the AGT data, but did not provide support for a separate general, bifactor that might instantiate the “goal complexes” posited by Elliot (2005) and Urdan and Kaplan (2020). Our findings do not necessarily imply the idea of goal complexes should be abandoned, but it may be the case that they must be modeled differently (e.g., latent profile analyses; Holden et al., 2021; Schwinger & Wilds, 2012; Tuominen et al., 2020; Wang et al., 2016). However, our findings do support the use of ESEM methods over CFA techniques, which call into question the results of latent profile analyses, or any other analyses, that derived factor scores via CFA.

We were particularly pleased to find, for each motivation theory, at least metric measurement invariance across semesters, suggesting substantial consistency across context, even when the semesters observed varied in course modality (i.e., in-person lectures in Fall 2019 and 2021, whereas Fall 2020 was delivered online). The clear superiority of the ESEM-based measurement models for expectancy value theory and achievement goal theory, and their consistency across semesters, argues for future investigations of optimal motivation measurement models not just to inform their conceptualization (Eccles & Wigfield, 2020; Wigfield & Koenka, 2020) but also to inform how scores from those measurement models are used to deploy interventions (Rosenzweig et al., 2022) and refine education practice (e.g., creating adaptive classroom goal structures, Canning et al., 2020; making value more salient, Priniski et al., 2018; reducing perceptions of cost, Rosenzweig et al., 2020). In essence, our findings suggest the motivation instruments show sufficient consistency to be used to inform practice, in contexts similar to the one we studied, but their scores are best derived using more advanced techniques (e.g., ESEM) than is typical in practice and prior research.

Our second research question involved an investigation of the predictive validity of motivation scores derived from our measurement models. After controlling for confounds due to SRL interventions we deployed, which were ancillary to this study, we found motivation factors, measured once at the beginning of the semester, predicted a small amount of the variance in postsecondary biology academic outcomes. However, even findings of small relations can have practical value, particularly when they have low financial, time, and effort costs to administer, such as is the case with these motivation surveys (Kraft, 2020). Initial student characteristics, such as motivation, prior knowledge, and learning behaviors, have been used to predict achievement with accuracy (Bernacki, Chavez, Usebeck, 2020). In turn, such predictions have been used to deliver interventions to support STEM learning and avert negative achievement and retention outcomes (Bernacki, Vosicka, et al., 2020; Bernacki et al., 2021). Similarly, prior research has shown the motivational profiles that students exhibit at the outset of a STEM course are predictive of the ways they engage with course content and the success they achieve (Hong et al., 2020), and that initial and evolving motivations have implications for persistence (Hong & Bernacki, 2022). Our findings suggest these early identification, prediction, and intervention efforts might be enhanced with more complex, yet more accurate, motivation measurement models. Such modeling has already proven useful regarding how to parse the state and trait aspects of EVT (Part et al, 2023) and we expect additional longitudinal research utilizing bifactor and ESEM modeling would prove equally insightful.

In the particular introductory biology course we focused upon in this study, across all three semesters, attainment value was a useful predictor of achievement. This finding aligns with research on the importance of attainment to STEM undergraduates' identity (Perez, et al., 2014) and how students' self-concept influences the type of STEM degree they pursue (Toh & Watt, 2022). Moreover, Part et al (2020) found that attainment value was of sufficient centrality that it tracked closely with the more general perception of a course's value for those who enroll in pursuit of a STEM degree and career. Our findings further bolster the argument for attainment value as an important aspect of EVT.

Also, the general factor for expectancy value theory, subjective task value, predicted achievement in each semester of this particular biology course. Subjective task value very well could be a critical predictor of performance and engagement in STEM (Hong & Bernacki, 2022). These findings provide further evidence that for continued evolution of EVT, there is potentially great value in specifying "exactly how the various components would aggregate to form either the STV of individual achievement-related task or the relative STVs across several different task(s)" (Eccles & Wigfield, 2020, p. 6). It may be that EVT should be expanded to formally conceptualize a general SVT factor, with subsequent implications for how to reconceptualize the more specific values and costs via a process of epistemic iteration (Greene, 2023). As students navigate changes in the instructional context during the course of their STEM degree pursuit, bifactor modeling that parses the specificity and dimensionality of student motivation may provide important insight into students' perception of their courses, the opportunities they provide for learning, and their expectancy that they can engage productively in them (Byrnes & Miller, 2007).

For research questions 1 and 2, nearly all our findings held across non-biology and biology majors in this course. The intriguing exceptions to this consistency were the relationships between motivation factors and academic outcomes in Fall 2020, which was during the COVID-19 pandemic and required moving the course to a completely online modality. In this semester, utility value, attainment value, and self-efficacy motivation factors were weaker predictors of achievement for non-biology majors, compared to biology majors. Overall, the r-squared values for the regression of academic outcomes on motivation factors, and the ancillary intervention, were small for non-biology majors, but medium to large for biology majors. These findings are post hoc and exploratory, therefore definitive explanations are not possible or warranted. Nonetheless, it is intriguing that during a time of high curricular and extracurricular stress (Tate & Warschauer, 2022), these relations between motivation and academic achievement differed depending upon a student's major. It may be that established effects of course modality on students' engagement (Martin & Borup, 2022) and sense of presence (Shea et al., 2022) must be expanded to include effects on motivation, particularly for those students for whom the course is not a major requirement. This is clearly an area for future research, hopefully via intentional and benevolent natural experiments on the effects of changes in course modality, rather than changes forced by unfortunate circumstances.

Limitations

Our analyses are necessarily restricted to those students who chose to participate in each phase of our study. Some participants did not complete the motivation survey and therefore could not be included in any analyses. In two of the three semesters, these participants were more likely to withdraw from the course than their peers who did complete the survey. In all semesters, participants who did not complete the motivation survey had lower final exam and course grades than their peers. Given this, the internal validity of our inferences from the motivation modeling and measurement invariance analyses must be necessarily circumscribed. More descriptive and qualitative research is needed to understand why participants decided against completing the motivation survey and about the nature of their motivation (Meyer & Schutz, 2020). Participants' reasons for choosing against completing the motivation survey may be driven by a factor, or factors, that also relate to their achievement scores. If this is the case, accounting for this factor may change the results of analyses of the motivation surveys' factor structure and measurement invariance. Likewise, some participants completed the motivation surveys but did not agree to participate in our RCT, thus limiting the internal validity of inferences from the results of our analyses for Research Question 2. When we examined whether consent status related to motivation scores, across the eleven motivation factors and three semesters (i.e., 33 analyses), only three statistically significant relationships were found, all in Fall 2019. Whether or not students chose to participate in the RCT was not statistically significantly related to withdrawal, final exam score, or course grade except in one semester (i.e., final exam grade in Fall 2019). These findings suggest consent status was not a factor in our analyses, except perhaps in Fall 2019. Nonetheless, future research is needed to understand these students' reticence to participate in the RCT, and whether it is related to differences in the nature or amount of their motivation.

There is value in studying motivation across consecutive semesters of the same course, but such value necessarily comes with limitations on the generalizability of the findings. There is a clear need for more research on motivation measurement models with different cohorts of students, in different courses and majors, and in different contexts. Indeed, it may be the case that the underlying conceptualizations of motivation in achievement goal theory and expectancy value theory do not adequately capture the phenomena for some groups of students (Kumar et al., 2018). As such, more re-imaged scholarship is needed to determine the breadth, depth, and proper scope of the conceptualizations of motivation that drive measurement modeling (Greene, 2023).

This study was nested within a larger scholarship project focused on self-regulated learning interventions. As such, the effects of these interventions had to be considered when examining how motivation scores predicted student achievement in the course. This confounding can be controlled statistically, but such adjustments do not account for the possibility that motivation interacted with the interventions, or the prediction modeling, to affect student outcomes. Ideally, future research would not involve such interventions, so the role of motivation in STEM achievement can be examined without significant confounds.

Implications for Practice

Our findings for research question 1, in particular the importance of considering more advanced ESEM and Bifactor-ESEM models when deriving EVT and AGT scores, present something of a dilemma for practitioners. It is likely unreasonable to expect most educators to be able to derive and test various measurement models, including particularly complex ESEM varieties, before making inferences about student motivation from scores from EVT, AGT, and self-efficacy instruments. Nonetheless, failure to account for a subjective task value factor (Part et al., 2020, 2023) or failure to use ESEM to derive factor scores may result in biased and potentially uninformative data on student motivation (Asparouhov et al., 2015). Such data may prove misleading, or at worst harmful, for supporting students' STEM performance and aspirations. Therefore, we see three clear implications for practice from our findings. First, given the growing evidence of subjective task value as an important aspect of EVT, practitioners should be made aware of this conceptualization of values and costs and scoring guides for EVT instruments should help practitioners derive both general and specific EVT scores. Relatedly, second, practitioners and researchers would benefit from a common, online source for pooling and analyzing motivation data. Anonymized scores from motivation instruments could be uploaded to a central repository where various kinds of ESEM and Bifactor-ESEM analyses could be automated and agglomerated. Such a tool could quickly produce motivation scores for practitioners while simultaneously building a large dataset of scores for further human-driven analyses that continued to confirm or refine the measurement, and perhaps even the situativity, of these constructs. Third, given our findings regarding the predictive value of utility value and subjective task value in particular, practitioners who lack the resources to administer and analyze motivation instrument data could focus their efforts on these constructs, mirroring successful interventions based on these phenomena (Rosenzweig et al., 2022), perhaps with added nuance given our consistent finding of a positively predictive, general subjective task value factor but the intriguingly negative relationship between outcomes and the specific utility value factor, in our context.

Conclusion

Many postsecondary STEM students benefit from motivational monitoring and support, which in turn should promote positive academic and career outcomes (Rosenzweig et al., 2022). To support these students, researchers and practitioners need instruments that produce useful and accurate indicators of student motivation. Prior research has suggested more complex measurement models (e.g., ESEM, Bifactor-ESEM models) produce more accurate motivation scores than more typical models (e.g., CFA; Hamilton et al., 2018; Lohbeck et al., 2022; Part et al., 2020, 2023), and our findings provided further support for such claims. In addition, we found these models were consistent across three semesters of students, with at least metric measurement invariance across expectancy value, achievement goal, and self-efficacy theory constructs. Scores from these models, in turn, predicted academic outcomes across all three semesters, particularly scores from those factors measuring attainment value and subjective task value. Researchers and practitioners wishing to monitor and intervene upon students' STEM motivations to promote academic and career success should consider using these more complex measurement models to derive the most accurate and useful indicators of which students would benefit from support and how best to help them. Finally, these findings suggest the need to further investigate refined conceptualizations of STEM student motivation.

References

Asparouhov, T., Muthén, B., & Morin, A. J. S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeyer et al. J*ournal of Management, 41,* 1561-1577. https://doi.org/10.1177/0149206315591075

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*(1), 1-26. https://doi.org/10.1146/annurev.psych.52.1.1

Barron, K. E., & Harackiewicz, J. M. (2001). Achievement goals and optimal motivation: testing multiple goal models. *Journal of Personality and Social Psychology*, *80*(5), 706. https://doi.org/10.1037/0022-3514.80.5.706

Battle, A., & Wigfield, A. (2003). College women's value orientations toward family, career, and graduate school. *Journal of Vocational Behavior*, *62*(1), 56-75. https://doi.org/10.1016/S0001-8791(02)00037-4

Bernacki, M. L. (2018). Examining the cyclical, loosely sequenced, and contingent features of self-regulated learning: Trace data and their analysis. In D. H. Schunk, & J. A. Greene (Eds.), *Handbook of Self-Regulated Learning and Performance* (pp. 370–387). Routledge. https://doi.org/10.4324/9781315697048-24

Bernacki, M. L., Chavez, M. M., & Uesbeck, P. M. (2020). Predicting STEM achievement with learning management system data: Prediction modeling and a test of an early warning system. *Computers & Education, 158*, 103999. https://doi.org/10.1016/j.compedu.2020.103999

Bernacki, M. L., Vosicka, L., & Utz, J. C. (2020). Can brief, web-delivered training help STEM undergraduates "learn to learn"*?. Journal of Educational Psychology, 112*(4), 765–781. https://doi.org/10.1037/edu0000405

Bernacki, M. L., Vosicka, L., Utz, J. C. & Warren, C. (2021). Effects of digital learning skill training on the academic performance of undergraduates in science and mathematics. *Journal of Educational Psychology, 113*(6), 1107–1125. https://doi.org/10.1037/edu0000485

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education, 27*, 1-13. https://doi.org/10.1016/j.iheduc.2015.04.007

Byrnes, J. P., & Miller, D. C. (2007). The relative importance of predictors of math and science achievement: An opportunity–propensity analysis. *Contemporary Educational Psychology, 32*(4), 599-629. https://doi.org/10.1016/j.cedpsych.2006.09.002

Cannady, M. A., Greenwald, R., & Harris, K. N. (2014). Problematizing the STEM pipeline metaphor: Is the STEM pipeline metaphor serving our students and the STEM workforce? *Science Education, 98*(3), 443-460. https://doi.org/10.1002/sce.21108

Canning, E. A., LaCosse, J., Kroeper, K. M., & Murphy, M. C. (2020). Feeling like an imposter: the effect of perceived classroom competition on the daily psychological experiences of first-generation college students. *Social Psychological and Personality Science, 11*(5), 647-657. https://doi.org/10.1177/1948550619882032

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J-P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality, 80*(1), 219-251. https://doi.org/10.1111/j.1467-6494.2011.00739.x

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*(2), 189-225. https://doi.org/10.1207/s15327906mbr4102_5

Chen, X. (2015). STEM attrition among high-performing college students: Scope and potential causes. *Journal of Technology and Science Education*, *5*(1), 41-59. http://dx.doi.org/10.3926/jotse.136

Church, M. A. et al., (2001). Perceptions of classroom environment, achievement goals, and achievement outcomes. *Journal of Educational Psychology, 93*(1), 43. https://doi.org/10.1037/0022-0663.93.1.43

Cogliano, M., Bernacki, M. L., & Kardash, C. M. (2020). A metacognitive retrieval practice intervention to improve undergraduates' monitoring and control processes and use of performance feedback for classroom learning. *Journal of Educational Psychology, 113*(7), 1421-1440. https://doi.org/10.1037/edu0000624

Cogliano, M. C., Bernacki, M.L., Hilpert J. C., & Strong, C. (2022). A self-regulated learning analytics prediction-and-intervention design: Detecting and supporting struggling biology students. *Journal of Educational Psychology, 114*(8), 1801-1816. https://doi.org/10.1037/edu0000745

Conley, A. M. (2012). Patterns of motivation beliefs: Combining achievement goal and expectancy-value perspectives. *Journal of Educational Psychology, 104*(1), 32–47. https://doi.org/10.1037/a0026042

Cromley, J. G., Perez, T., & Kaplan, A. (2016). Undergraduate stem achievement and retention: Cognitive, motivational, and institutional factors and solutions. *Policy Insights from the Behavioral and Brain Sciences, 3*(1), 4–11. https://doi.org/10.1177/2372732215622648

Deci, E. L., & Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*, 227–268. http://dx.doi.org/10.1207/S15327965PLI1104_01

Dompnier, B., Darnon, C., & Butera, F. (2013). When performance-approach goals predict academic achievement and when they do not: A social value approach. *British Journal of Social Psychology, 52*(3), 587–596. https://doi.org/10.1111/bjso.12025

Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., & Meece, J. L. (1983). Expectations, values, and academic behaviors. In J. T. Spence (Ed.). *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75–146). San Francisco, CA: Freeman.

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Eddy, S. L., & Hogan, K. A. (2014). Getting under the hood: How and for whom does increasing course structure work?. *CBE—Life Sciences Education*, *13*(3), 453-468. https://doi.org/10.1187/cbe.14-03-0050

Elliot, A. J. (2005). A conceptual history of the achievement goal construct. In A. Elliot, & C. Dweck (Eds.). *Handbook of competence and motivation* (pp. 52–72). Guilford Press.

Elliot, A. J., & Dweck, C. S. (2005). Competence and motivation. In A. J. Elliot & C. S. Dweck (Eds.) *Handbook of competence and motivation* (pp. 3-12). Guildford Publications.

Elliot, A. J., & McGregor, H. A. (2001). A 2× 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*(3), 501-519. https://doi.org/10/1037//0022-3514.80.3.501

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology, 100*(3), 613–628. https://doi.org/10.1037/0022-0663.100.3.613

Enders, C. K. (2022). *Applied Missing Data Analysis (2nd Edition).* The Guilford Press.

Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology, 41*, 232–244. https://doi.org/10.1016/j.cedpsych.2015.03.002

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74-78. https://doi.org/10.1016/j.paid.2016.06.069

Greene, J. A. (2018). *Self-regulation in education.* Routledge.

Greene, J. A. (2022). What can educational psychology learn from, and contribute to, scholarship on theory development? *Educational Psychology Review, 34*, 3011-3035. https://doi.org/10.1007/s10648-022-09682-5 https://doi.org/10.1007/s10648-022-09682-5

Gunnell, K. E., & Gaudreau, P. (2015). Testing a bi-factor model to disentangle general and specific factors of motivation in self-determination theory. *Personality and Individual Differences*, *81*, 35-40. https://doi.org/10.1016/j.paid.2014.12.059

Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology, 94*(3), 638–645. https://doi.org/10.1037/0022-0663.94.3.638

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta- analyses relating to achievement.* Routledge.

Holden, S. M., Mueller, C. E., Harrell-Williams, L. M., Ford, J. M., & Jones, M. H. (2021). Comparison of motivational latent profiles using the PALS and AGQ-R. *Contemporary Educational Psychology*, *67*, 101999. https://doi.org/10.1016/j.cedpsych.2021.101999

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2*(1), 41-54. https://doi.org/10.1007/BF02287965

Hong, W., & Bernacki, M. L. (2022). Initial and evolving perceptions of value and cost of engaging in undergraduate science course work and effects on achievement and persistence. *Journal of Educational Psychology, 114*(5), 1005-1027. https://doi.org/10.1037/edu0000717

Hong, W., Bernacki, M. L., & Perera, H. N. (2020). A latent profile analysis of undergraduates' achievement motivations and metacognitive behaviors, and their relations to achievement in science. *Journal of Educational Psychology, 112*(7), 1409-1430. https://doi.org/10.1037/edu0000445

Howard, J. L., Gagné, M., Morin, A. J. S., & Forest, J. (2018). Using bifactor exploratory structural equation modeling to test for a continuum structure of motivation. J*ournal of Management, 44(*7), 2638-2664. Https://doi.org/10.1177/0149206316645653

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Jiang, S., Simpkins, S. D., & Eccles, J. S. (2020). Individuals' math and science motivation and their subsequent STEM choices and achievement in high school and college: A longitudinal study of gender and college generation status differences. *Developmental Psychology*, *56*(11), 2137. https://doi.org/10.1037/dev0001110

Kaplan, A., & Maehr, M. L. (2007). The contributions and prospects of goal orientation theory. *Educational Psychology Review*, *19*(2), 141-184. https://doi.org/10.1007/s10648- 006-9012-5

Karabenick, S. A. (2004). Perceived achievement goal structure and college student help seeking. *Journal of Educational Psychology, 96*(3), 569. https://doi.org/10.1037/0022-0663.96.3.569

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford Publications.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241-253. https://doi.org/10.3102/0013189X20912798

Krou, M. R., Fong, C. J., & Hoff, M. A. (2021). Achievement motivation and academic dishonesty: A meta-analytic investigation. *Educational Psychology Review, 33*, 427-458. https://doi.org/10.1007/s10648-020-09557-7

Kumar, R., Zusho, A., & Bondie, R. (2018). Weaving cultural relevance and achievement motivation into inclusive classroom cultures. *Educational Psychologist*, *53*(2), 78-96. https://doi.org/10.1080/00461520.2018.1432361

Lee, S. Y., Friedman, S., Christiaans, E., & Robinson, K. A. (2022). Valuable but costly? University students' expectancy-value-cost profiles in introductory chemistry courses. *Contemporary Educational Psychology*, *69*, 102056. https://doi.org/10.1016/j.cedpsych.2022.102056

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198-1202. https://doi.org/10.1080/01621459.1988.10478722

Lohbeck, A., Tóth-Király, I., & Morin, A. J. (2022). Disentangling the associations of academic motivation with self-concept and academic achievement using the bifactor exploratory structural equation modeling framework. *Contemporary Educational Psychology*, *69*, 102069. https://doi.org/10.1016/j.cedpsych.2022.102069

Martin, F., & Borup, J. (2022). Online learner engagement: Conceptual definition, research themes, and supportive practices. *Educational Psychologist, 57*(3), 162-177. https://doi.org/10.1080/00461520.2022.2089147

Meyer, D. K., & Schutz, P. A. (2020). Why talk about qualitative and mixed methods in educational psychology? Introduction to special issue. *Educational Psychologist*, *55*(4), 193-196. https://doi.org/10.1080/00461520.2020.1796671

Morin, A. J. S., Marsh, H., & Nagengast, B. (2013). Exploratory structural equation modeling: An introduction. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed.): 395-436. IAP.

Morin, A.J.S., Myers, N.D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling (ESEM) and bifactor-ESEM. In G. Tenenbaum, & Eklund, R.C. (Eds.), Handbook of Sport Psychology, 4th Edition. London, UK. Wiley. https://doi.org/10.1002/9781119568124.ch51

Muenks, K., Miller, J. E., Schuetze, B. A., & Whittaker, T. A. (2023). Is cost separate from or part of subjective task value? An empirical examination of expectancy-value versus expectancy-value-cost perspectives. *Contemporary Educational Psychology*, 102149. https://doi.org/10.1016/j.cedpsych.2023.102149

Muthén, L.K. and Muthén, B.O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén.

Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K. T., & Trautwein, U. (2011). Who took the "×" out of expectancy-value theory?: A psychological mystery, a substantive methodological synergy, and a cross-national generalization. *Psychological Science, 22*(8), 1058–1066. https://doi.org/10.1177/0956797611415540

National Science Board. (2021). Digital products and federal policy for the innovation economy. In *Advancing Commercialization of Digital Products from Federal Laboratories*. National Academies Press (US).

Olson, S., & Riordan, D. G. (2012). Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Report to the President. *Executive Office of the President*.

Pajares, F., Britner, S. L., & Valiante, G. (2000). Relation between achievement goals and self-beliefs of middle school students in writing and science. *Contemporary Educational Psychology, 25*(4), 406-422. https://doi.org/10.1006/ceps.1999.1027

Part, R., Perera, H. N., Marchand, G. C., & Bernacki, M. L. (2020). Revisiting the dimensionality of subjective task value: Towards clarification of competing perspectives. *Contemporary Educational Psychology*, *62*, 101875. https://doi.org/10.1016/j.cedpsych.2020.101875

Part, R., Perera, H. N., Mefferd, K, & Miller, C. J. (2023). Decomposing trait and statevariability in general and specific subjective task value beliefs. *Contemporary Educational Psychology, 72, 102112*. https://doi.org/10.1016/j.cedpsych.2022.102112

Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology, 101*(1), 115. https://doi.org/10.1037/a0013383

Perez, T., Cromley, J. G., & Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *Journal of Educational Psychology, 106*(1), 315–329. https://doi.org/10.1037/a0034027

Perez, T., Wormington, S. V., Barger, M. M., Schwartz-Bloom, R. D., Lee, Y.-K., & Linnenbrink-Garcia, L. (2019). Science expectancy, value, and cost profiles and their proximal and distal relations to undergraduate science, technology, engineering, and math persistence. *Science Education, 103*(2), 264–286. https://doi.org/10.1002/sce.21490

Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology, 92*(3), 544. https://doi.org/10.1016/B978-012109890-2/50043-3

Priniski, S. J., Hecht, C. A., & Harackiewicz, J. M. (2018). Making learning personally meaningful: A new framework for relevance research. *The Journal of Experimental Education, 86*(1), 11-29. https://doi.org/10.1080/00220973.2017.1380589

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*, 667-696. https://doi.org/10.1080/00273171.2012.715555

Robinson, K. A., Lee, Y., Bovee, E. A., Perez, T., Walton, S. P., Briedis, D., & Linnenbrink-Garcia, L. (2019). Motivation in transition: Development and roles of expectancy, task values, and costs in early college engineering. *Journal of Educational Psychology, 111*(6), 1081–1102. https://doi.org/10.1037/edu0000331

Robinson, K. A., Lira, A. K., Walton, S. P., Briedis, D., & Linnenbrink-Garcia, L. (2022). Instructional Supports for Motivation Trajectories in Introductory College Engineering. *AERA Open*, *8*, 23328584221083662. https://doi.org/10.1177/23328584221083662

Rosenzweig, E. Q., Harackiewicz, J. M., Hecht, C. A., Priniski, S. J., Canning, E. A., Tibbetts, Y., Asher, M. W., & Hyde, J. S. (2021). College students' reasons for leaving biomedical fields: Disenchantment with biomedicine or attraction to other fields? *Journal of Educational Psychology, 113*(2), 351–369. https://doi.org/10.1037/edu0000456

Rosenzweig, E. Q., Wigfield, A., & Eccles, J. S. (2022). Beyond utility value interventions: The why, when, and how for next steps in expectancy-value intervention research. *Educational Psychologist*, *57*(1), 11-30. https://doi.org/10.1080/00461520.2021.1984242

Rosenzweig, E. Q., Wigfield, A., & Hulleman, C. S. (2020). More useful or not so bad? Examining the effects of utility value and cost reduction interventions in college physics. *Journal of Educational Psychology, 112*(1), 166. https://doi.org/10.1037/edu0000370

Schwinger, M., & Wild, E. (2012). Prevalence, stability, and functionality of achievement goal profiles in mathematics from third to seventh grade. *Contemporary Educational Psychology*, *37*(1), 1-13. https://doi.org/10.1016/j.cedpsych.2011.08.001

Senko, C., Hulleman, C. S., & Harackiewicz, J. M. (2011). Achievement goal theory at the crossroads: Old controversies, current challenges, and new directions. *Educational Psychologist, 46*(1), 26–47. https://doi.org/10.1080/00461520.2011.538646

Senko, C., & Tropiano, K. L. (2016). Comparing three models of achievement goals: Goal orientations, goal standards, and goal complexes. *Journal of Educational Psychology*, *108*(8), 1178. https://doi.org/10.1037/edu0000114

Shea, P., Richardson, J., & Swan, K. (2022). Building bridges to advance the community of inquiry framework for online learning. *Educational Psychologist, 57*(3), 148-161. https://doi.org/10.1080/00461520.2022.2089989

Simon, R. A., Aulls, M. W., Dedic, H., Hubbard, K., & Hall, N. C. (2015). Exploring student persistence in STEM programs: a motivational model. *Canadian Journal of Education*, *38*(1), 1-27. https://www.jstor.org/stable/10.2307/canajeducrevucan.38.1.09

Sommet, N., & Elliot, A. J. (2017). Achievement goals, reasons for goal pursuit, and achievement goal complexes as predictors of beneficial outcomes: Is the influence of goals reducible to reasons?. *Journal of Educational Psychology, 109*(8), 1141-1162. https://doi.org/10.1037/edu0000199

Tas, Y., & Tekkaya, C. (2010). Personal and contextual factors associated with students' cheating in science. *The Journal of Experimental Education, 78*(4), 440-463. https://doi.org/10.1080/00220970903548046

Tate, T., & Warschauer, M. (2022). Equity in online learning. *Educational Psychologist, 57*(3), 192-206. https://doi.org/10.1080/00461520.2022.2062597

Theobald, M. (2021). Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: A meta-analysis. *Contemporary Educational Psychology*, *66*, 101976. https://doi.org/10.1016/j.cedpsych.2021.101976

Toh, L., & Watt, H. M. (2022). How do adolescent mathematical self-concept and values explain attainment of different kinds of STEM degrees in adulthood?. *Contemporary Educational Psychology, 69*, 102057. https://doi.org/10.1016/j.cedpsych.2022.102057

Tuominen, H., Niemivirta, M., Lonka, K., & Salmela-Aro, K. (2020). Motivation across a transition: Changes in achievement goal orientations and academic well-being from elementary to secondary school. *Learning and Individual Differences*, *79*, 101854. https://doi.org/10.1016/j.lindif.2020.101854

Urdan, T., & Kaplan, A. (2020). The origins, evolution, and future directions of achievement goal theory. *Contemporary Educational Psychology*, *61*, 101862. https://doi.org/10.1016/j.cedpsych.2020.101862

Van Yperen, N. W., Elliot, A. J., & Anseel, F. (2009). The influence of mastery-avoidance goals on performance improvement. *European Journal of Social Psychology*, *39*(6), 932-943. https://doi.org/10.1002/ejsp.590

Wang, J. C., Morin, A. J., Liu, W. C., & Chian, L. K. (2016). Predicting physical activity intention and behaviour using achievement goal theory: A person-centred analysis. *Psychology of Sport and Exercise*, *23*, 13-20. https://doi.org/10.1016/j.psychsport.2015.10.004

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68-81. https://doi.org/10.1006/ceps.1999.1015

Wigfield, A., & Koenka, A. C. (2020). Where do we go from here in academic motivation theory and research? Some reflections and recommendations for future work. *Contemporary Educational Psychology*, *61*, 101872. https://doi.org/10.1016/j.cedpsych.2020.101872

Wormington, S. V., & Linnenbrink-Garcia, L. (2017). A new look at multiple goal pursuit: The promise of a person-centered approach. *Educational Psychology Review*, *29*(3), 407-445. https://doi.org/10.1007/s10648-016-9358-2

Yeh, Y. C., Kwok, O. M., Chien, H. Y., Sweany, N. W., Baek, E., & McIntosh, W. A. (2019). How College Students' Achievement Goal Orientations Predict Their Expected Online Learning Outcome: The Mediation Roles of Self-Regulated Learning Strategies and Supportive Online Learning Behaviors. *Online Learning*, *23*(4), 23-41. https://doi.org/10.24059/olj.v23i4.2076

Zheng, L. (2016). The effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: A meta-analysis. *Asia Pacific Education Review, 17*, 187–202. https://dx.doi.org/10.1007/s12564-016-9426-9

Table 1

Course population, withdraw count, consent status, and analytic sample sizes, by semester

Semester	Fall 2019	Fall 2020	Fall 2021
Course Population	488	432	549
Withdrew With No Motivation Survey Data	1	9	9
Did Not Withdraw But No Motivation Survey Data	78	51	22
Analytic Sample for Research Question 1	409 (16% missing)	372 (14% missing)	518 (9.4% missing)
Did Not Consent to Participate in RCT, Did Not Withdraw	66	55	103
Did Not Complete Pretest	6	3	1
Course Subsample Available for Research Question 2 Analyses	337	314	414
Withdrew After Completing Motivation Survey	11 (3.3% attrition)	9 (2.9% attrition)	2 (.5% attrition)
Analytic Sample Size for Research Question 2	326 (33% missing from course population)	305 (29% missing from course population)	412 (25% missing from course population
Total Withdrawal Percentage for Entire Course Population	2.5%	4.2%	2.0%

Note: In Fall 2021, one section of the course did not participate in the randomized controlled trial; they are included in the “Did not Consent” column.

Table 2

Group Size and Demographics by Semester

	Fall 2019			Fall 2020			Fall 2021		
	Total	RQ1	RQ2	Total	RQ1	RQ2	Total	RQ1	RQ2
n	488	409	332	432	372	308	549	518	412
Experimental Group									
Excluded from experiment	96	77		69	62		99	93	
Non-Flagged Control	200	163	163						
Flagged Control	66	60	60						
Flagged Advice Page	126	109	109						
Randomly Assigned Control				269	223	222			
Randomly Assigned Science of Learning Lite				94	87	86			
Non-Flagged Biology Multimedia							207	199	198
Flagged Biology Multimedia							74	69	67
Flagged Science of Learning Multimedia							74	68	64
Flagged Coaching							95	89	83
Academic Groups									
Biology major	24.0%	25.2%	25.3%	25.5%	24.2%	25.5%	27.3%	26.1%	27.9%
Honors program	6.3%	5.9%	6.3%	4.5%	5.4%	4.5%	5.2%	5.6%	5.3%
Demographic Groups									
Female	69.9%	71.7%	70.2%	67.3%	66.9%	66.0%	69.2%	68.7%	68.0%
First generation college student	23.7%	24.3%	24.3%	25.9%	25.5%	24.6%	20.6%	21.2%	19.3%
Under-represented minorities in STEM fields	23.8%	22.3%	22.8%	25.4%	25.8%	25.4%	29.8%	30.0%	29.5%
Ethnicity									
Caucasian	68.9%	68.9%	68.7%	66.0%	65.9%	66.8%	62.4%	63.5%	64.3%
Asian	13.7%	13.9%	14.8%	18.5%	18.8%	19.0%	18.0%	18.3%	18.4%
African American	12.7%	11.7%	12.3%	13.0%	13.2%	12.9%	16.0%	16.2%	16.3%
Hispanic or Latinx	9.2%	8.6%	8.4%	11.8%	12.1%	12.3%	12.6%	12.9%	12.9%
Non-Hispanic Latinx	5.1%	4.9%	4.5%	5.1%	5.1%	4.2%	6.2%	6.4%	5.8%
Native American	2.7%	2.4%	2.7%	2.5%	2.4%	2.9%	3.1%	3.3%	2.9%
Hawaiian or other Pacific Islander	0.4%	0.5%	0.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Not Specified	3.1%	3.2%	2.7%	1.4%	1.3%	0.6%	1.6%	1.5%	1.5%

Table 3

Descriptive Statistics

	Fall 2019		Fall 2020		Fall 2021	
Assessment or Item	Mean	SD	Mean	SD	Mean	SD
Final exam	75.950	14.219	75.807	13.044	76.271	12.578
Final grade in course	82.712	9.737	83.982	10.856	84.279	8.974
Effort Cost 1	1.950	1.127	2.060	1.124	1.93	1.085
Effort Cost 2	2.380	1.211	2.500	1.279	2.40	1.185
Effort Cost 3	2.200	1.308	2.220	1.231	2.17	1.218
Effort Cost 4	2.170	1.240	2.150	1.175	2.06	1.047
Opportunity Cost 1	2.470	1.376	2.580	1.423	2.71	1.469
Opportunity Cost 2	2.140	1.243	2.030	1.165	2.19	1.213
Opportunity Cost 3	1.920	1.151	1.830	1.101	1.86	1.101
Opportunity Cost 4	1.920	1.159	1.900	1.147	1.90	1.107
Psychological Cost 1	4.120	1.425	4.320	1.326	4.40	1.330
Psychological Cost 2	4.070	1.346	4.050	1.384	4.16	1.366
Psychological Cost 3	3.470	1.440	3.630	1.394	3.49	1.403
Psychological Cost 4	3.570	1.467	3.520	1.417	3.41	1.498
Attainment Value 1	5.300	.792	5.140	.845	5.14	.795
Attainment Value 2	5.240	.727	5.140	.780	5.19	.772
Attainment Value 3	5.710	.562	5.630	.612	5.69	.560
Attainment Value 4	5.170	.873	5.100	.909	5.18	.916
Intrinsic Value 1	4.460	1.019	4.460	1.007	4.45	.976
Intrinsic Value 2	4.540	.967	4.630	.976	4.68	.838
Intrinsic Value 3	4.440	.966	4.480	.953	4.58	.895
Intrinsic Value 4	4.730	.907	4.780	.866	4.83	.830
Utility Value 1	5.160	1.148	4.950	1.334	5.10	1.132
Utility Value 2	4.250	1.112	4.170	1.211	4.17	1.111
Utility Value 3	5.110	1.007	4.980	1.131	5.08	1.049
Utility Value 4	5.270	1.046	5.080	1.175	5.24	1.064
Self-efficacy 1	4.680	.893	4.740	.902	4.71	.877
Self-efficacy 2	4.610	.899	4.690	.900	4.68	.889
Self-efficacy 3	5.260	.794	5.270	.752	5.28	.756
Self-efficacy 4	5.170	.800	5.140	.808	5.15	.775
Self-efficacy 5	5.100	.827	5.070	.874	5.12	.855
Mastery Approach Goals 1	6.450	.859	6.290	.966	6.37	.867
Mastery Approach Goals 2	6.680	.694	6.490	.839	6.59	.697
Mastery Approach Goals 3	6.620	.725	6.520	.765	6.53	.750
Performance Approach Goals 1	6.100	1.146	6.060	1.136	6.07	1.151
Performance Approach Goals 2	5.950	1.255	6.000	1.168	6.07	1.145
Performance Approach Goals 3	4.980	1.506	5.070	1.462	5.12	1.487
Performance Avoidance Goals 1	5.470	1.633	5.500	1.570	5.74	1.477
Performance Avoidance Goals 2	5.330	1.645	5.43	1.603	5.63	1.493
Performance Avoidance Goals 3	5.290	1.609	5.33	1.595	5.51	1.504

Table 4

Data-Model Fit and Measurement Invariance Testing Across Semesters

Model	Chi-square value/df	AIC	BIC	SABIC	RMSEA (90% confidence interval)	CFI	SRMR	Chi-square difference test with less restrictive model
Expectancy Value Theory (Values and Cost only)								
CFA Configural	2604.214(711)***	95069.035	97736.376	96135.320	.074 (.071, .077)	.881	.080	n/a
CFA Metric	2603.081(747)***	95051.002	97527.816	96041.123	.071 (.068, .074)	.883	.095	34.747(36)
CFA Scalar	2705.489(783)***	95078.217	97364.507	95992.175	.071 (.068, .074)	.879	.089	99.412(36)***
Bifactor-CFA Configural [a]	2669.942(684)***	95124.474	97934.705	96247.881	.077 (.074, .080)	.875	.148	n/a
Bifactor-CFA Metric [a]	2680.620(766)***	95090.490	97466.749	96040.414	.071 (.069, .074)	.879	.154	86.753(82)
Bifactor-CFA Scalar [a]	2775.636(800)***	95115.470	97311.790	95993.463	.071 (.068, .074)	.876	.139	92.235(34)***
ESEM Configural	1316.898(441)***	75137.669	77882.594	76195.864	.068 (.064, .072)	.944	.026	n/a

ESEM Metric	1521.319(657)***	74995.950	76624.296	75623.693	.055 (.052, .059)	.944	.042	235.022(216)
ESEM Scalar	1634.701(693)***	75038.736	76480.984	75594.737	.056 (.053, .060)	.940	.046	118.767(36)***
Bifactor-ESEM Configural	803.925(387)***	74663.080	77687.150	75828.888	.050 (.045, .055)	*.973*	*.017*	n/a
Bifactor-ESEM Metric	1042.684(625)***	74508.143	76301.908	75199.657	.039 (.035, .043)	*.973*	.037	258.327(238)
Bifactor-ESEM Scalar	1126.775(659)***	*74528.494*	*76146.501*	*75152.251*	*.040 (.036, .044)*	.970	.040	87.708(34)***
				Achievement Goal Theory				
CFA Configural	276.225(72)***	49248.333	50296.216	49667.231	.076 (.067, .086)	.943	.274	n/a
CFA Metric	277.396(84)***	49239.066	50223.441	49632.576	.069 (.060, .078)	.946	.413	9.123(12)
CFA Scalar	299.636(96)***	49233.020	50153.886	49601.142	.066 (.057, .074)	.943	.125	17.617(12)
Bifactor-CFA Configural [a b]	153.894(57)***	49108.239	50235.507	49558.871	.059 (.048, .070)	.973	.237	n/a
Bifactor-CFA Metric [a b]	170.946(85)***	49107.917	50086.999	49499.311	.045 (.036, .055)	.976	.467	30.573(28)

Bifactor-CFA Scalar [a b]	185.319(95)***	49100.185	50026.344	49470.423	.044 (.035, .053)	.975	.243	12.298(10)
ESEM Configural	62.412(36)**	30061.459	30712.797	30312.556	.041 (.023, .058)	.992	*.012*	n/a
ESEM Metric	95.889(72)*	30037.534	30502.775	30216.889	.028 (.009, .041)	*.993*	.038	34.110(36)
ESEM Scalar	111.224(84)*	*30028.278*	*30431.487*	*30183.719*	*.027 (.010, .040)*	.992	.042	15.107(12)
Bifactor-ESEM Configural [c]								
Bifactor-ESEM Metric [c]								
Bifactor-ESEM Scalar [c]								
Self-Efficacy								
CFA Configural	260.310(15)***	31423.048	32042.252	31670.579	.183 (.164, .203)	.912	.487	n/a
CFA Metric	281.373(23)***	31412.569	31989.434	31643.174	.151 (.136, .168)	.907	.194	4.470(8)
CFA Scalar	307.416(31)***	31406.688	31941.214	31620.368	.135 (.121, .149)	.901	.302	9.937(8)
CFA with one residual	31.859(12)**	31108.104	31743.185	31361.987	.058 (.034, .083)	*.993*	*.128*	n/a

covariance Configural								
CFA with one residual covariance Metric	42.474(20)**	31100.694	31693.435	31337.646	.048 (.028, .068)	.992	.373	8.139(8)
CFA with one residual covariance Scalar	52.450(28)**	*31093.139*	*31643.543*	*31313.166*	*.042 (.024, .060)*	.991	.341	8.457(8)

*p < .05; **p < .01; ***p < .001
Note: Values in *italics* are optimal on that criterion, for those motivation factors.
[a] Model's latent variable covariance matrix was not positive definite
[b] To achieve model convergence, the residual variance of one indicator was set to zero.
[c] Model did not terminate normally, numerous estimation errors.

Table 5

Final Factor Analytic Model Standardized Loadings For Best-Fitting Expectancy-Value Theory, Achievement Goal Theory, and Self-Efficacy Measurement Models

Factor	Item	Fall 2019	Fall 2020	Fall 2021
Effort Cost	Effort Cost 1	.619***	.581***	.559***
	Effort Cost 2	.610***	.559***	.566***
	Effort Cost 3	.477***	.486***	.470***
	Effort Cost 4	.553***	.547***	.571***
	Opportunity Cost 1	.019	.018	.016
	Opportunity Cost 2	.061**	.062**	.056**
	Opportunity Cost 3	.167***	.177***	.165***
	Opportunity Cost 4	.164***	.158***	.155***
	Psychological Cost 1	.108*	.109*	.103*
	Psychological Cost 2	.100*	.094*	.090*
	Psychological Cost 3	-.106**	-.104**	-.098**
	Psychological Cost 4	-.209***	-.208***	-.185***
	Attainment Value 1	-.179***	-.161***	-.167***
	Attainment Value 2	-.007	-.007	-.006
	Attainment Value 3	-.124**	-.113**	-.110**
	Attainment Value 4	.017	.016	.014
	Intrinsic Value 1	.017	.017	.017
	Intrinsic Value 2	-.019	-.019	-.019
	Intrinsic Value 3	.063*	.059*	.061*
	Intrinsic Value 4	.039	.037	.037
	Utility Value 1	-.067*	-.056*	-.060*
	Utility Value 2	.018	.016	.017
	Utility Value 3	-.069*	-.060*	-.063*
	Utility Value 4	-.057	-.047	-.049
Opportunity Cost	Effort 1	.115***	.113***	.109***
	Effort 2	.118***	.112***	.115***
	Effort 3	.155***	.164***	.160***
	Effort 4	.100***	.103**	.109**
	Opportunity Cost 1	.702***	.680***	.631***
	Opportunity Cost 2	.815***	.855***	.778***
	Opportunity Cost 3	.649***	.714***	.671***
	Opportunity Cost 4	.675***	.678***	.670***
	Psychological Cost 1	.184***	.193***	.183***
	Psychological Cost 2	.180***	.177***	.169***
	Psychological Cost 3	.011	.011	.011
	Psychological Cost 4	-.049	-.051	-.046
	Attainment Value 1	.008	.008	.008
	Attainment Value 2	.049	.047	.045
	Attainment Value 3	-.047	-.047	-.044
	Attainment Value 4	.080*	.076*	.071*

	Intrinsic Value 1	.058	.058	.059
	Intrinsic Value 2	.037	.039	.040
	Intrinsic Value 3	.089**	.087**	.091**
	Intrinsic Value 4	.050	.050	.050
	Utility Value 1	.146***	.126***	.138***
	Utility Value 2	.098**	.090**	.094**
	Utility Value 3	.125***	.112***	.119***
	Utility Value 4	.121***	.103***	.108***
Psychological Cost	Effort 1	-.128***	-.116***	-.120***
	Effort 2	-.033	-.029	-.032
	Effort 3	.042	.041	.043
	Effort 4	.014	.014	.015
	Opportunity Cost 1	.218***	.196***	.194***
	Opportunity Cost 2	.106**	.103**	.100**
	Opportunity Cost 3	.044	.045	.046
	Opportunity Cost 4	.057	.053	.056
	Psychological Cost 1	.823***	.806***	.814***
	Psychological Cost 2	.813***	.745***	.761***
	Psychological Cost 3	.582***	.549***	.559***
	Psychological Cost 4	.618***	.596***	.571***
	Attainment Value 1	.093**	.081**	.091**
	Attainment Value 2	.119**	.107**	.110**
	Attainment Value 3	.286***	.254***	.266***
	Attainment Value 4	.147***	.130**	.130**
	Intrinsic Value 1	-.008	-.004	-.008
	Intrinsic Value 2	-.001	-.001	-.001
	Intrinsic Value 3	.002	.002	.002
	Intrinsic Value 4	.012	.011	.012
	Utility Value 1	.103***	.083***	.096***
	Utility Value 2	.017	.014	.016
	Utility Value 3	.114***	.095***	.108***
	Utility Value 4	.097**	.077**	.086**
Attainment Value	Effort 1	-.037	-.038	-.038
	Effort 2	-.050*	-.051*	-.052*
	Effort 3	.033	.037	.036
	Effort 4	.027	.030	.032
	Opportunity Cost 1	-.017	-.017	-.016
	Opportunity Cost 2	.036*	.040**	.036**
	Opportunity Cost 3	.019	.022	.021
	Opportunity Cost 4	.025	.027	.027
	Psychological Cost 1	-.031	-.034	-.032
	Psychological Cost 2	-.005	-.006	-.005
	Psychological Cost 3	.072**	.077**	.074**
	Psychological Cost 4	.105***	.115***	.105***
	Attainment Value 1	.169***	.168***	.177***
	Attainment Value 2	.655***	.667***	.648***

	Attainment Value 3	.298***	.300***	.297***
	Attainment Value 4	.796***	.801***	.758***
	Intrinsic Value 1	.058*	.062*	.063*
	Intrinsic Value 2	.039	.044	.045
	Intrinsic Value 3	.042	.044	.047
	Intrinsic Value 4	.036	.038	.038
	Utility Value 1	.042	.038	.042
	Utility Value 2	.075**	.073**	.077**
	Utility Value 3	.061**	.058**	.062**
	Utility Value 4	.194***	.175***	.185***
Intrinsic Value	Effort 1	.085***	.082**	.081**
	Effort 2	-.003	-.003	-.003
	Effort 3	.009	.010	.010
	Effort 4	-.009	-.009	-.010
	Opportunity Cost 1	.010	.010	.009
	Opportunity Cost 2	.042*	.044*	.030*
	Opportunity Cost 3	.063**	.069*	.066**
	Opportunity Cost 4	.028	.028	.028
	Psychological Cost 1	-.103*	-.108*	-.104*
	Psychological Cost 2	-.076	-.074	-.072
	Psychological Cost 3	.061	.061	.059
	Psychological Cost 4	.158***	.163***	.148***
	Attainment Value 1	.363***	.340***	.360***
	Attainment Value 2	.114**	.109**	.106**
	Attainment Value 3	.111**	.105**	.105**
	Attainment Value 4	.032	.030	.029
	Intrinsic Value 1	.645***	.652***	.667***
	Intrinsic Value 2	.640***	.674***	.692***
	Intrinsic Value 3	.729***	.714***	.755***
	Intrinsic Value 4	.653***	.648***	.655***
	Utility Value 1	.083*	.072*	.079*
	Utility Value 2	347***	.318***	.336***
	Utility Value 3	.185***	.166***	.178***
	Utility Value 4	.024	.021	.022
Utility Value	Effort 1	-.202***	-.217***	-.221***
	Effort 2	-.102***	-.107***	-.115***
	Effort 3	.081**	.094**	.096**
	Effort 4	.038	.043	.048
	Opportunity Cost 1	.089***	.095***	.092***
	Opportunity Cost 2	.098***	.113***	.107***
	Opportunity Cost 3	.068**	.083**	.081**
	Opportunity Cost 4	.073***	.081***	.084***
	Psychological Cost 1	-.032	-.037	-.037
	Psychological Cost 2	-.039	-.042	-.043
	Psychological Cost 3	.119***	.133***	.133***
	Psychological Cost 4	.172***	.195***	.185***

	Attainment Value 1	.236***	.244***	.267***
	Attainment Value 2	.161***	.170**	.172***
	Attainment Value 3	147***	.154**	.159***
	Attainment Value 4	.271***	.283***	.278***
	Intrinsic Value 1	.166***	.185***	.195***
	Intrinsic Value 2	.193***	.228***	.242***
	Intrinsic Value 3	.136***	.147**	.161**
	Intrinsic Value 4	.163***	.178***	.187***
	Utility Value 1	.753***	.716***	.817***
	Utility Value 2	.406***	.410***	.448***
	Utility Value 3	.688***	.680***	.759***
	Utility Value 4	.680***	.637***	.700**
Subjective Task Value[a]	Effort 1	.552***	.538***	.551***
	Effort 2	.491***	.468***	.504***
	Effort 3	.538***	.568***	.585***
	Effort 4	.618***	.635***	.705***
	Opportunity Cost 1	.481***	.466***	.455***
	Opportunity Cost 2	.511***	.535***	.513***
	Opportunity Cost 3	.450***	.494***	.490***
	Opportunity Cost 4	.469***	.470***	.490***
	Psychological Cost 1	.028	.029	.029
	Psychological Cost 2	.065	.064	.064
	Psychological Cost 3	.561***	.569***	.572***
	Psychological Cost 4	.609***	.630***	.596***
	Attainment Value 1	-.430***	-.403***	-.445***
	Attainment Value 2	-.395***	-.380***	-.386***
	Attainment Value 3	-.278***	-.264***	-.274***
	Attainment Value 4	-.399***	-.379***	-.375***
	Intrinsic Value 1	-.463***	-.468***	-.499***
	Intrinsic Value 2	-.501***	-.527***	-.564***
	Intrinsic Value 3	-.527***	-.516***	-.569***
	Intrinsic Value 4	-.508***	-.504***	-.532***
	Utility Value 1	-.449***	-.387***	-.445***
	Utility Value 2	-.286***	-.263***	-.289***
	Utility Value 3	-.442***	-.396***	-.445***
	Utility Value 4	-.453***	-.385***	-.426***
Self-Efficacy	Self-efficacy 1	.570***	.570***	.570***
	Self-efficacy 2	.615***	.615***	.615***
	Self-efficacy 3	.570***	.570***	.570***
	Self-efficacy 4	.692***	.692***	.692***
	Self-efficacy 5	.740***	.740***	.740***
Mastery Approach	Mastery Approach 1	.682***	.699***	.633***
	Mastery Approach 2	.918***	.886***	.850***
	Mastery Approach 3	.786***	.843***	.730***
	Performance Approach 1	.042*	.049*	.040*
	Performance Approach 2	.002	.002	.002

	Performance Approach 3	-.005	-.005	-.005
	Performance Avoidance 1	-.037*	-.044*	-.039*
	Performance Avoidance 2	.027**	.032**	.028**
	Performance Avoidance 3	.020	.023	.020
Performance Approach	Mastery Approach 1	.091*	.081*	.087*
	Mastery Approach 2	-.090**	-.076**	-.086**
	Mastery Approach 3	.041	.038	.039
	Performance Approach 1	.985***	1.001***	.961***
	Performance Approach 2	.856***	.910***	.890***
	Performance Approach 3	.536***	.551***	.551***
	Performance Avoidance 1	.147***	.152***	.158***
	Performance Avoidance 2	-.043	-.044	-.047
	Performance Avoidance 3	-.012	-.012	-.012
Performance Avoidance	Mastery Approach 1	.028	.025	.026
	Mastery Approach 2	.028	.024	.026
	Mastery Approach 3	-.031	-.029	-.029
	Performance Approach 1	-.158***	-.162	-.149***
	Performance Approach 2	-.059*	-.063*	.059*
	Performance Approach 3	.259***	.268***	.258***
	Performance Avoidance 1	.639***	.666***	.669***
	Performance Avoidance 2	.927***	.955***	.968***
	Performance Avoidance 3	.900***	.924***	.895***

* $p < .05$, ** $p < .01$, *** $p < .001$

[a] The Bifactor-ESEM solution for this factor reversed all factor loadings, therefore lower scores should be interpreted as higher subjective task value, and higher scores indicating lower subjective task value

Note: Despite metric measurement invariance for EVT models across semesters and scalar measurement invariance for AGT models across semesters, standardized factor loadings vary across semesters because of differences in latent factor variances.

Table 6

Expectancy Value Theory, Bifactor-ESEM Metric Invariance Latent Factor Correlations

	Effort	Opportunity	Psychological	Attainment	Intrinsic	Utility	Subjective Task Value [a]
Effort	.683 / .654 / .652	-.108	.011	-.019	-.097	-.023	-.107
Opportunity	-.073	.826 / .850 / .802	.182*	.081	.001	-.076	-.033
Psychological	.063	.015	.845 / .811 / .818	-.110	-.010	.031	-.012
Attainment	-.046	.008	-.091	.729 / .737 / .696	-.022	-.002	-.007
Intrinsic	-.138	.106	.025	.011	.720 / .787 / .807	-.111	.136
Utility	.038	.111	.141	-.158	-.016	.728 / .683 / .837	.030
Subjective Task Value [a]	.035	-.188*	-.059	.052	.045	-.215*	.873 / .873 / .885

* $p < .05$, ** $p < .01$, *** $p < .001$

[a] The Bifactor-ESEM solution for this factor reversed all factor loadings, therefore lower scores should be interpreted as higher subjective task value, and higher scores indicating lower subjective task value

Note: Fall 2019 latent factor correlations rotated to be zero; Fall 2020 correlations on lower triangle, Fall 2021 on upper triangle; Coefficient H reliability values, shown as Fall 2019 / Fall 2020 / Fall 2021, on diagonal

Table 7

Achievement Goal Theory ESEM Scalar Invariance Latent Factor Correlations

	Mastery Approach	Performance Approach	Performance Avoidance
Mastery Approach	.887 / .876 / .815		
Performance Approach	.454*** / .389*** / .336***	.973 / 1.000 / .942	
Performance Avoidance	.202** / .134* / .045	.576*** / .658*** / .679***	.918 / .945 / .952

* $p < .05$, ** $p < .01$, *** $p < .001$

Note: Correlations shown as Fall 2019 / Fall 2020 / Fall 2021; Coefficient H reliability values, shown as Fall 2019 / Fall 2020 / Fall 2021, on diagonal

Table 8

Correlations Among Motivation and Outcome Variables for Research Question 2, Fall 2019

	Effort	Opportunity	Psychological	Attainment	Intrinsic	Utility	Subjective Task Value [a]	Self-Efficacy	Mastery Approach	Performance Approach	Performance Avoidance	Final Exam
Effort												
Opportunity	.039											
Psychological	-.041	-.024										
Attainment	.062	.019	.036									
Intrinsic	.055	.042	.043	-.074								
Utility	-.009	.066	.025	.027	.032							
Subjective Task Value [a]	.030	.014	.066	-.001	-.069	-.067						
Self-Efficacy	-.041	-.044	-.031	.136*	.178**	.014	-.578**					
Mastery Approach	-.122*	-.067	.098	.113*	.087	.143**	-.386**	.450**				
Performance Approach	.042	.011	.153**	.143**	-.033	.096	-.160**	.204**	.526**			
Performance Avoidance	.163**	.038	.245**	.144**	-.058	.100	-.052	.105	.226**	.617**		
Final Exam	.053	.008	.085	.086	-.025	-.073	-.150**	.125*	.096	.040	.068	
Course Grade	.007	.046	-.006	.179**	-.057	-.031	-.172**	.114*	.096	-.035	-.043	.886**

* $p < .05$, ** $p < .01$, *** $p < .001$

[a] The Bifactor-ESEM solution for this factor reversed all factor loadings, therefore lower scores should be interpreted as higher subjective task value, and higher scores indicating lower subjective task value

Table 9

Correlations Among Motivation and Outcome Variables for Research Question 2, Fall 2020

	Effort	Opportunity	Psychological	Attainment	Intrinsic	Utility	Subjective Task Value [a]	Self-Efficacy	Mastery Approach	Performance Approach	Performance Avoidance	Final Exam
Effort												
Opportunity	-.061											
Psychological	.011	.046										
Attainment	-.020	.021	-.085									
Intrinsic	-.095	.170**	.068	.023								
Utility	.021	.209**	.249**	-.178**	.054							
Subjective Task Value [a]	.136*	-.138*	-.014	.004	-.011	-.357**						
Self-Efficacy	-.038	.058	-.079	.068	.141*	.193**	-.494**					
Mastery Approach	-.244**	.215**	.219**	.089	.305**	.445**	-.503**	.320**				
Performance Approach	-.046	.130*	.270**	.042	.046	.207**	-.196**	.192**	.425**			
Performance Avoidance	.109	.019	.262**	.003	-.019	.104	-.070	.049	.139*	.704**		
Final Exam	-.091	.083	-.033	.086	-.011	.014	-.194**	.059	.059	-.007	-.023	
Course Grade	-.051	.111	.030	-.003	.026	.046	-.146*	.038	.057	-.003	-.028	.837**

* $p < .05$, ** $p < .01$, *** $p < .001$

[a] The Bifactor-ESEM solution for this factor reversed all factor loadings, therefore lower scores should be interpreted as higher subjective task value, and higher scores indicating lower subjective task value

Table 10

Correlations Among Motivation and Outcome Variables for Research Question 2, Fall 2021

	Effort	Opportunity	Psychological	Attainment	Intrinsic	Utility	Subjective Task Value [a]	Self-Efficacy	Mastery Approach	Performance Approach	Performance Avoidance	Final Exam
Effort												
Opportunity	-.093											
Psychological	-.030	.226**										
Attainment	.008	.111*	-.089									
Intrinsic	-.054	.041	.052									
Utility	-.012	-.004	.068	.015	-.085							
Subjective Task Value [a]	-.069	.016	.013	-.091	.120*	-.015						
Self-Efficacy	.059	.007	-.151**	.162**	.080	.074	-.521**					
Mastery Approach	-.060	.111*	.091	.317**	.152**	.255**	-.378**	.376**				
Performance Approach	.017	.158**	.254**	.186**	.105*	.065	-.162**	.208**	.413**			
Performance Avoidance	.016	.084	.252**	.079	.074	-.053	-.022	.068	.052	.726**		
Final Exam	.056	.142**	.129**	.099*	-.037	.010	-.158**	.049	.084	.072	-.022	
Course Grade	.062	.125*	.143**	.094	-.037	.037	-.202**	.088	.145**	.105*	-.001	.852**

* $p < .05$, ** $p < .01$, *** $p < .001$

[a] The Bifactor-ESEM solution for this factor reversed all factor loadings, therefore lower scores should be interpreted as higher subjective task value, and higher scores indicating lower subjective task value

Table 11

Final Exam Score Regressed on Latent Factors, by Semester

Latent Factor	Fall 2019 Standardized Regression Coefficient	Fall 2020 Standardized Regression Coefficient	Fall 2021 Standardized Regression Coefficient
Non-Biology Major *n* / Biology Major *n*	242/84	227/78	297/115
Effort Cost	.044/.045	-.073/-.077	.033/.035
Opportunity Cost	.025/.026	.067/.070	.104*/.107*
Psychological Cost	.075/.094	-.001/-.001	.086/.093
Attainment Value	.093/.103	**.154/-.140*	.101*/.085*
Intrinsic Value	-.042/-.045	-.009/-.009	-.001/-.001
Utility Value	-.095/-.066	*.032/-.203**	.019/.012
Subjective Task Value [a]	-.137*/-.135*	-.198/-.189	-.153**/-.168**
Self-efficacy	-.020/.022	-.029/-.028	-.056/-.052
Mastery Approach	.064/.037	-.079/-.035	-.046/-.035
Performance Approach	-.060/-.062	-.027/-.030	.096/.103
Performance Avoidance	-.034/.045	.002/.002	-.121/-.130
Non-flagged Group [b]	.116/.145		
Advice Page Group [b]	-.127*/-.210*		
Science of Learning Lite Group [c]		-.002/-.003	
Non-Flagged Biology Multimedia Group[d]			.317***/.337***
Science of Learning Multimedia Group[d]			-.020/-.020
Coaching Group[d]			-.014/.014
Model R^2	.132***/147**	.060*/.143*	.190***/.182***
Sum of Squared Standardized Regression Coefficients for Statistically Significant Motivation Latent Factors Only	.019/.018	.024/.041	.041/.047
Model chi-square(df), p-value	32.801(26), .168	23.799(19), .204	23.409(28), .712

* $p < .05$, ** $p < .01$, *** $p < .001$

Note: Standardized path coefficients and R^2 values shown as Non-Biology Major / Biology Major. They differ due to different variances in outcome variables across biology major status groups. Path coefficients that statistically significantly differ from one another are shown in italics and are found only in Fall 2020.

[a] The Bifactor-ESEM solution for this factor reversed all factor loadings, therefore lower scores should be interpreted as higher subjective task value, and higher scores indicating lower subjective task value

[b] Comparison group in Fall 2019 was flagged, control students

[c] Comparison group in Fall 2020 was control students

[d] Comparison group in Fall 2021 was flagged, control students

Table 12

Course Grade Regressed on Latent Factors, by Semester

Latent Factor	Fall 2019 Standardized Path Coefficient	Fall 2020 Standardized Path Coefficient	Fall 2021 Standardized Path Coefficient
Non-Biology Major *n* / Biology Major *n*	242/84	227/78	297/115
Effort Cost	.027/.029	-.083/-.078	.041/.044
Opportunity Cost	.063/.069	.108/.099	.081/.085
Psychological Cost	.060/.081	.012/.012	.093/.101
Attainment Value	.135**/.159**	*.072/-.180*	.075/.064
Intrinsic Value	-.042/-.048	.013/.010	.000/.000
Utility Value	-.102*/-.077	*.055/-.158*	.032/.019
Subjective Task Value [a]	-.110/-.116	-.193*/-.162*	-.162**/-.181**
Self-efficacy	-.009/.010	*-.093/.213*	-.042/-.040
Mastery Approach	.131/.082	-.079/-.030	.015/.012
Performance Approach	-.119/-.133	.001/.001	.082/.089
Performance Avoidance	-.002/-.003	-.018/.-.017	-.094/-.103
Non-flagged Group[b]	.158*/.212*		
Advice Page Group[b]	-.127/-.172		
Science of Learning Lite Group [c]		.025/.024	
Biology Multimedia Group [d]			.345**/.374**
Science of Learning Multimedia Group [d]			-.017/-.020
Coaching Group [d]			.006-.014
Model R^2	.145***/.166**	.054*/.199*	.209***/.213***
Sum of Squared Standardized Regression Coefficients for Statistically Significant Motivation Latent Factors Only	.029/.025	.037/.026	.026/.033
Model chi-square(df), p-value	32.801(26), .168	23.799(19), .204	23.409(28), .712

* $p < .05$, ** $p < .01$, *** $p < .001$

Note: Standardized path coefficients and R^2 values shown as Non-Biology Major / Biology Major. They differ due to different variances in outcome variables across biology major status groups. Path coefficients that statistically significantly differ from one another are shown in italics and are found only in Fall 2020.

[a] The Bifactor-ESEM solution for this factor reversed all factor loadings, therefore lower scores should be interpreted as higher subjective task value, and higher scores indicating lower subjective task value

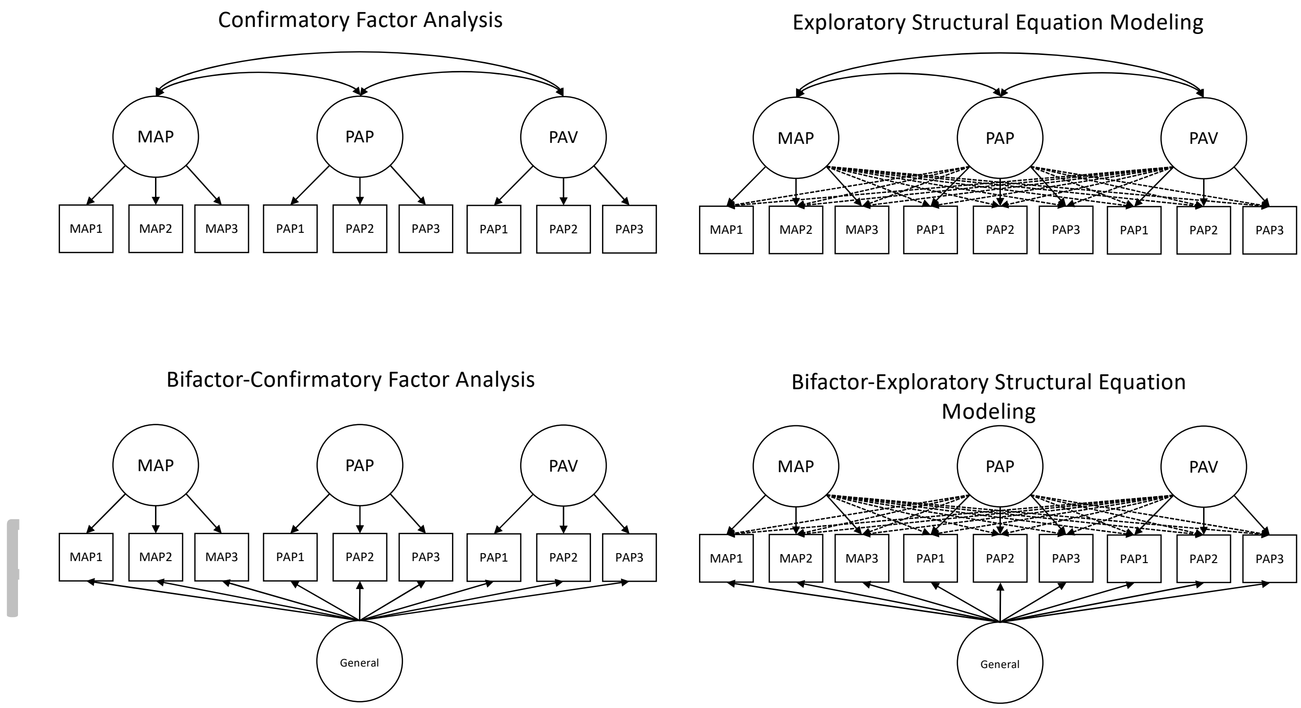
[b] Comparison group in Fall 2019 was flagged, control students

[c] Comparison group in Fall 2020 was control students

[d] Comparison group in Fall 2021 was flagged, control students

Figure 1

Methods of Modeling Motivation Specific and General Factors



Note: Models shown reflecting various ways of modeling item responses to a measure of achievement goal theory. Models for other motivation theories, such as expectancy value theory, would look similar, with more or fewer specific factors, as warranted by the theory.
MAP = mastery approach achievement goal; PAP = performance approach achievement goal;
PAV = performance avoidance achievement goal