10-18-2017

A Motivational Account of the Undergraduate Experience in Science: Brief Measures of Students' Self-System Appraisals, Engagement in Coursework, and Identity as a Scientist

Emily Saxton
Portland State University

Cailin Currie
Portland State University

Ellen A. Skinner
Portland State University, ellen.skinner@pdx.edu

Gwen Shusterman
Portland State University

A Motivational Account of the Undergraduate Experience in Science: Brief Measures of Students' Self-system Appraisals, Engagement in Coursework, and Identity as a Scientist

Ellen Skinner, Emily Saxton, Cailin Currie, and Gwen Shusterman

Department of Psychology Department of Chemistry

STEM Institute

Portland State University

Corresponding author:

Ellen A. Skinner
Psychology Department
Portland State University
PO Box 751
Portland OR 97207-0751

Telephone: (503)725-3966
e-mail: ellen.skinner@pdx.edu

This research was supported in part by an award to Portland State University under the Howard Hughes Medical Institute Science Education Program.

A Motivational Account of the Undergraduate Experience in Science: Brief Measures of Students' Self-system Appraisals, Engagement in Coursework, and Identity as a Scientist

As part of long-standing efforts to promote undergraduates' success in science, researchers have investigated the instructional strategies and motivational factors that promote student learning and persistence in science coursework and majors. This study aimed to create a set of brief measures that educators and researchers can use as tools to examine the undergraduate motivational experience in science classes. To identify key motivational processes, we drew on Self-Determination Theory (SDT), which holds that students have fundamental needs-- to feel competent, related, and autonomous-- that fuel their intrinsic motivation. When educational experiences meet these needs, students engage more energetically and learn more, cumulatively contributing to a positive identity as a scientist. Based on information provided by 1013 students from 8 classes in biology, chemistry, and physics, we constructed conceptually-focused and psychometrically-sound survey measures of three sets of motivational factors: (1) students' appraisals of their own *competence*, *autonomy*, and *relatedness*; (2) the quality of students' behavioral and emotional *engagement* in academic work; and (3) students' emerging *identities as scientists*, including their *science identity*, *purpose in science*, and *science career plans*. Using an iterative confirmatory process, we tested short item sets for unidimensionality and internal consistency, and then cross-validated them. Tests of measurement invariance showed that scales were generally comparable across disciplines. Most importantly, scales and final course grades showed correlations consistent with predictions from SDT. These measures may provide a window on the student motivational experience for educators, researchers, and interventionists who aim to improve the quality of undergraduate science teaching and learning.

A Motivational Account of the Undergraduate Experience in Science: Brief Measures of Students' Self-system Appraisals, Engagement in Coursework, and Identity as a Scientist

Over the last several decades, undergraduate institutions have become increasingly focused on their role in ensuring their students' persistence and success in science, technology, engineering, and mathematics (STEM) coursework and majors (American Society for Engineering Education, 2009, 2012; Hawwash, 2007; King, 2008; National Academy of Engineering, 2004; President's Council of Advisors on Science and Technology, 2012). Discipline-based educational researchers have helped academic departments identify a range of student-centered pedagogical strategies that foster the kinds of active learning found to promote students' mastery of complex STEM knowledge (Singer & Smith, 2013). Research suggests that these new ways of teaching and learning are especially important to the success and persistence of students from ethnic and racial minority groups and from low socioeconomic backgrounds, as well as women and first generation students, who otherwise are underrepresented in STEM majors and careers (Espinosa, 2011; Hurtado, Newman, Tran, & Chang, 2010; Tsui, 2007).

As part of this research, studies have begun to point to the importance of "non-cognitive" or "affective" factors (Singer, Nielsen, & Schweingruber, 2012). These factors, which also seem to predict college success and persistence more generally (Liu, Bridgemen, & Adler, 2012; Robbins et al., 2004), involve a wide range of student psychosocial dispositions, such as feelings of self-efficacy (Adedokun, Bessenbacher, Parker, Kirkham, & Burgess, 2013), intrinsic motivation (Glynn & Koballa, 2006), and a positive science identity (e.g., Chang, Eagan, Lin, & Hurtado, 2011). Such factors may provide a motivational advantage, helping students stay committed to the hard work that high performance in STEM coursework demands. In fact, studies suggest that some of the pathways through which pedagogical strategies promote student

STEM learning likely involve activating these motivational factors—for example, by counteracting societal stereotypes about the academic aptitude of specific groups or by highlighting the relevance of STEM careers to solving societal problems (e.g., Gasiewski, Eagan, Garcia, Hurtado, & Chang, 2012).

As evidence about the role of motivational factors accumulates, the field has become increasingly interested in incorporating measures of these attributes in research and interventions to enhance student success (Gasiewski et al., 2012; Glynn, & Koballa, 2006; Singer et al., 2012). Such assessments, which offer a window into the student motivational experience in STEM classes, can serve at least three important functions. First, they are valuable to ***educators***, because they provide actionable information about students' predispositions and attitudes. Such information can feed into ongoing pedagogical decisions, and can also alert instructors to individual students who need extra support (Glynn & Koballa, 2006). Second, assessments of motivational factors are helpful to ***researchers***, especially those interested in creating process-oriented accounts of the pathways to STEM persistence and success. Such studies can trace the steps from pedagogy to performance, examining, for example, whether student engagement is a necessary condition for high quality learning (e.g., Olson & Riordan, 2012) or whether the development of a strong science identity is a prerequisite to persistence for students from underrepresented minority groups (e.g., Carlone & Johnson, 2007). Third, motivational assessments are useful to ***interventionists*** because they provide benchmarks to chart the progress of programs designed to improve STEM teaching and learning. Moreover, if motivational factors are key to the student experience in STEM, then theories and research on motivation may suggest additional pedagogical and interpersonal strategies that can be incorporated by both educators and interventionists (Zusho, Pintrich, & Coppola, 2003).

Models of Student Motivational and Identity Development

Two main approaches have been used to identify relevant motivational processes and map them using quantitative surveys. First, discipline-based educational researchers have used bottom-up qualitative strategies (such as focus groups) to get a sense of students' experiences in the classroom; they have then generated pools of items to capture these experiences, and relied on exploratory factor analyses (EFA) to help distinguish clusters of connected items, and label them according to existing motivational constructs (e.g., Glynn, Taasoobshirazi, & Brickman, 2009). A second strategy, used by motivational experts, has been to apply collected wisdom from the larger field (based on many decades of research on K-12 students) in order to extend and adapt key constructs and measures for use with college students in STEM fields (e.g., Zusho et al., 2003). Both of these approaches have been useful in nominating multiple candidate processes as important to student success in STEM classes.

To date, however, most motivational measures have focused on social cognitive or value-expectancy models of motivation, which are centered on student self-efficacy as a pivotal motivational asset (see also Dalgety, Coll, & Jones, 2003). Around the pivot of self-efficacy, existing motivational measures fan out to cover different sets of constructs. For example, Zusho et al. (2003) included task value, mastery and performance goal orientations, interest, and anxiety, as well as a set of cognitive strategies of self-regulated learning, including rehearsal, organization, elaboration, and metacognitive self-regulation. In contrast, Glynn, Brickman, Armstrong, & Taasoobshirazi (2011) included intrinsic, career, and grade motivation in their measure, as well as a construct they call "self-determination" that included items such as "I study hard to learn science" and "I put enough effort into learning science," which self-determination theorists might instead label as "behavioral engagement" (Reeve, 2012; Skinner, Kindermann, Connell, & Wellborn, 2009b). Although self-efficacy has clearly been shown to be a strong predictor of engagement and performance in college STEM courses (e.g., Adedokun et al., 2013), more complex models emerging from the field of motivation suggest that additional factors may be in play.

Self-Determination Theory. In the current study, discipline-based educational researchers and motivational experts working together drew on Self-Determination Theory (SDT; Connell & Wellborn, 1991; Deci & Ryan, 1985, 2000; Reeve, 2012; Ryan & Deci, 2016, 2017) to identify the elements essential to motivation in STEM classrooms. SDT differs from social cognitive and expectancy-value models, which tend to view motivation as something that students have acquired as a result of prior socialization, such as the pattern of contingencies between students' past efforts and their performances, the values and goals espoused by parents, or the study strategies students have been taught (Schunk & DiBenedetto, 2016; Wigfield, Tonks, & Klauda, 2016). These previous socialization experiences, which may differ for students from disadvantaged backgrounds, are seen as the sources of students' current motivation — typically operationalized as expectancies (or efficacy), values, and goal orientations — which in turn, contribute to their goal-directed effort and persistence.

In contrast, SDT highlights the vital role of *intrinsic* motivation, common to all students regardless of background or history. This perspective is anchored by the assumption that students come with fundamental psychological needs, intrinsic to all humans, whose fulfillment provides the motivational "fire" that fuels engagement in learning. SDT focuses on three needs: (1) *competence*--the need to feel efficacious and capable; (2) *autonomy*--the need to experience one's true self as the source of motivation and action; and (3) *relatedness*--the need to connect deeply with others and to belong. Based on this assumption, which differs fundamentally from expectancy-value and social cognitive models, SDT can provide an alternative account of key intrinsic motivational processes that are upstream from success in STEM coursework, highlighting student experiences, self-system appraisals, engagement, and identity as contributors to learning and persistence [blinded reference].

Self-system appraisals. When students' intrinsic needs are met in STEM courses, this transforms students' experiences of STEM, their engagement in science, and eventually their own identities as STEM learners. This transformation can be charted empirically by assessing students' self-system processes, engagement, and identities as scientists (see Figure 1 and Table 1 for an overview). When needs for *competence* are met, students feel able to successfully complete demanding coursework and report high levels of perceived competence or self-efficacy; if this need is not met, students can feel discouraged and helpless. When needs for *autonomy* are met, students feel a sense of ownership for their own work and report high levels of personal commitment to learning; if this need is not met, students' can feel pressured, resentful, and adrift. When needs for *relatedness* are met, students feel at home with their classmates and in their classes and majors, and report high levels of belonging and connection; if this need is not met, students feel isolated and excluded, and are likely to look elsewhere for classes and majors where they feel more welcome.

Insert Table 1 and Figure 1 about here

Student engagement. Although the student experiences embodied in the self-system processes of competence, autonomy, and relatedness are not directly visible to instructors in the classroom, their effects are. These positive experiences have been shown to underlie the behaviors of students whom instructors view as "highly motivated" (Lee & Reeve, 2012). Many decades of research with adolescents and young adults have demonstrated that students who feel competent, autonomous, and related are also more likely to work hard, take initiative, follow-through, and persist on challenging assignments; they take advantage of enrichment opportunities (like review sessions, office hours, and tutoring); and they show interest, enthusiasm, and zest for learning (Christenson, Reschly &Wylie, 2012; Handelsman Briggs, Sullivan, & Towler, 2005; Wigfield et al., 2015). Together, these student actions are known collectively as *engagement* and they can be contrasted with *disaffection*, which is evident in students who are passive, reactive, discouraged, resentful, or who give up easily (Skinner et al., 2009b). Instructors correctly assume that, compared to students who are disaffected, engaged students will learn more, persist longer, and perform better in STEM coursework and majors (Handelsman et al., 2005). In fact, some studies suggest that enthusiastic engagement with academic material is a necessary condition for deep learning (see Reeve, 2012, for a review).

Identity as a scientist. Students' active engagement, persistence, and success in STEM classes should also, over time, cement a valuable internal motivational resource, namely, a strong *identity as a scientist*—which combines a personal science identity with future plans for a career involving science and a sense that science serves important societal purposes. A *science identity* reflects a student's deeply rooted conviction that he or she belongs in the world of science, endorsing a robust sense of himself or herself as "the kind of person" who resonates at a fundamental level with the core values and pursuits of the community of science. If students do not develop this strong motivational anchor, they may become more vulnerable to disaffection and desistance, especially in the face of academic or personal challenges and setbacks. *Science career plans* refer to the extent to which students see science as an integral part of their vocational aspirations. If students begin to doubt whether science will play a role in their futures, they may lose their resolve to persist in science coursework and majors. Finally, a *sense of purpose*, or certainty that STEM professions can contribute to the solution of important problems facing the world today, strengthens students' convictions that classwork and careers in STEM are meaningful, important, and worthwhile. Without a strong sense of purpose, students can begin to see STEM as meaningless and empty, and so not worth the effort to master. Studies suggest that feelings of competence, relatedness, and autonomy can contribute to a positive identity as a scientist, which in turn prepares students for future success in STEM careers or graduate education (e.g., Bauer, 2005). A strong identity as a scientist may be especially important for women, underrepresented minorities, and first generation students, as a resource when they encounter obstacles and discrimination (Carlone & Johnson, 2007; Chang et al., 2011; Lee, Alston, & Kahn, 2015).

Purpose of the Current Study

The purpose of the present study was to help create a window into undergraduates' motivational experiences in science classes, by developing a suite of conceptually-focused and psychometrically-sound motivational surveys for use by instructors, researchers, and interventionists. We created brief measures of key motivational factors and tested their psychometric functioning for use across three science disciplines: biology, chemistry, and physics. The three kinds of motivational factors were those identified by SDT as central to intrinsic motivation, persistence, and success, namely, self-systems (of competence, autonomy, and relatedness), engagement (behavioral and emotional engagement and disaffection), and identity as a scientist (including science identity, science career plans, and a sense of purpose in science).

In addition to providing an expanded range of constructs, we attempted to build on the work of other researchers who have created motivational assessments (e.g., Glynn et al., 2011) in three ways. First, we used a theory-driven approach in which we relied on SDT to help us target important aspects of students' classroom experiences, highlighting aspects that are not currently targeted in other motivational surveys, and that are otherwise invisible to instructors. These assessments represent core motivational constructs that have been shown to be active ingredients in promoting persistence, learning, and academic success, not just in STEM undergraduate classes, but across the spectrum of coursework and student groups. Second, to be useful to ***educators***, we attempted to construct assessments that were both practical and credible. Hence, measures were brief, so they could easily be incorporated into regular course activities; and items employed plain language high in face validity, so instructors could translate students' responses into improvements in the classroom. Third, to be useful to ***researchers*** and ***interventionists***, we attempted to construct assessments that were psychometrically sound and valid indicators of target constructs. Hence, we used confirmatory methods to test these brief theoretically-derived assessments for unidimensionality, internal consistency, cross-time stability, and invariance across disciplines. To test for predictive validity, we examined the extent to which these measures (1) showed the pattern of interrelations with one another hypothesized by SDT (as depicted in Figure 1) and (2) demonstrated clear connections with actual performance, with the expectation that students' reports of their motivational experiences would predict their actual final course grades.

Method

Sample and Design

Participants were undergraduates enrolled in eight science courses (in biology, chemistry, and physics) whose instructors were taking part in a longitudinal study of science pedagogy at an urban university in the Pacific Northwest. At Time 1 (T1) 856 students participated, and at Time 2 (T2) 574 students participated, with approximately 49% (n = 417) participating at both time points. Gender composition and ethnic background appear in Table 2.

Insert Table 2 about here

Procedures

At the beginning and end of Fall term 2015, instructors invited students to participate in online surveys and posted survey links on their course websites. A few points of extra credit were offered at each instructor's discretion; however, credit was awarded based on opening the link, and did not require survey completion. The T1 survey launched at the end of week 1 and remained open for two weeks. The T2 survey launched in week 9 of the 11-week term and remained open for two weeks, closing at the end of finals. Students responded to 83 survey items both times, in addition to demographic questions. Average completion time was 14 minutes. This larger item pool was used to derive the brief scales described below.

Motivational Scales

Because no existing SDT measures tapped these constructs in the science domain, items were adapted from standard motivational measures or adopted from previous pilot studies of undergraduate students [blinded citation]. Each of these measures has a history of psychometric and structural analyses verifying its dimensionality, so the primary goal for this study was to confirm that each measure, adopted for use in this new domain and/or age group, functioned well as an internally consistent unidimensional assessment across college students from three science disciplines.

Generation of items. In addition to our conceptual focus on SDT, we used two other strategies to create items. First, consistent with other STEM researchers (e.g., Glynn et al., 2011), we sought to create items that were conceptually clear but stated in language that was straightforward and jargon-free, language that participants themselves might use to describe their experiences (DeVellis, 2016). For example, in contrast to measures of self-efficacy that utilize a more complex sentence structure, such as, "I am not confident about understanding difficult science concepts," we generated more strongly-worded items, such as "I don't have the intelligence/brains to succeed in science." Items and scales so composed may have both empirical and practical advantages. Empirically, the clarity and incisiveness of such items may make it more likely to obtain satisfactory internal consistencies with only four to five items (Glynn et al., 2011). Practically, the high face validity of such items, as well as the brevity of the scales, recommends them to instructors as useful. Moreover, because surveys are brief and items correspond closely to undergraduates' lived experiences, students may be likely to respond more honestly and completely.

The second strategy for creating scales that were credible and useful was designed to deal with constructs that previous studies have shown are multi-dimensional. For these constructs, we generated items from multiple sub-dimensions and/or valences. For example, scales tapping engagement and disaffection include sub-dimensions like behavioral engagement in class and outside of class (Chi, 2013); and many of the scales include negative items that tap experiences that are the opposite of target constructs (e.g., a sense of *in*competence or feelings of *ex*clusion). This approach distinguishes the current scales from other brief measures of motivation, which often delete negative items from subscales as measurement development efforts proceed (e.g., Glynn et al., 2011, who removed markers of anxiety from subscales tapping self-efficacy). Although it can diminish unidimensionality or internal consistency, these kinds of scales may have conceptual, practical, and empirical advantages-- in that they provide coverage of a wider range of construct space (and student experiences) and so can show closer connections with target outcomes.

Initial item pools. Table 3 summarizes the source measures from which items were identified or adapted, along with item examples. Using the procedures just described, additional items were also created to fill out the item pools for each subscale.

Insert Table 3 about here

Self-system processes. We based our assessments on the foundation of previous measurement development work by SDT theorists, who have examined the multi-dimensionality of Competence, Autonomy, and Relatedness. However, in order for the current measures to be useful to educators and interventionists, we wanted go beyond the typical multi-dimensional assessments comprised of dozens of items tapping each self-system, to create very brief measures of 5 items each. For *Competence*, similar to other measures of perceived competence developed by SDT theorists (e.g., Skinner, Wellborn, & Connell, 1990), we generated 12 items tapping students' beliefs about their ability to succeed in science classes and the field of science and careers. These items mapped onto several aspects of Competence, including perceived control (e.g., "Even if they are challenging, I can do well in my science classes"), ability capacity (e.g., "I am good at science"), and unknown control (e.g., "When I do poorly in a science course, I usually can't figure out why," reverse-coded). Hence, the primary question was whether 5 items could be identified that spanned this rich conceptual space while still forming a unidimensional internally consistent subscale.

The item pool for *Autonomy* orientation entailed 6 items tapping students' personal commitment to the work in STEM classes and careers. Consistent with standard measures of autonomy that focus on the "why" of student action (e.g., Ryan & Connell, 1989), items were framed by the stem "Why do I do my classwork and homework for this course?". Students rated three underlying reasons for their participation: (a) identified reasons, or personal goals of learning (e.g., "Because I want to understand the subject"), (b) intrinsic reasons, based on inherent enjoyment of the activity (e.g., "Because it's fun to answer challenging science questions"), and (c) amotivation, which indicates little commitment (e.g., "For this class, I just learn the stuff I have to in order to pass the test(s)"). Because these three reasons represent different aspects of autonomy, the primary question was whether 5 items could be identified that not only covered this complex conceptual space, but also comprised a unidimensional internally consistent subscale.

Following other measures developed by SDT theorists (e.g., Chi, 2013), the item pool for sense of *Relatedness* consisted of 12 items tapping the extent to which students felt welcome and accepted in class and as science majors more generally. Items tapped students' sense of belonging in science classes (e.g., "This course is a good place for students like me"), with science students (e.g., "I fit in well with the other students in this class"), and in the science major more generally (e.g., "I'm not really sure that science is the right major for me," reverse-coded). Hence, the primary question was whether items covering this conceptual space could be identified that would form a unidimensional and internally consistent subscale.

Engagement versus disaffection. Structural analyses of domain-general measures of engagement versus disaffection in college students (Chi, 2013) and in youth (e.g., Skinner, Kindermann, & Furrer, 2009a) have distinguished four aspects of the construct: behavioral and emotional features of both engagement and disaffection. Hence, item pools were created or adapted for each dimension. The item pool for *Behavioral Engagement* comprised 9 items tapping students' effort and active participation in coursework both in class (e.g., "I pay attention in class") and outside of class (e.g., "I keep up with the work for this class"). The item pool for *Emotional Engagement* included 9 items tapping students' motivated emotions while participating in academic work, both inside class (e.g., "I enjoy the time I spend in this class") and outside of class (e.g., "The readings for this class are interesting"). The item pool for *Behavioral Disaffection* entailed 8 items tapping students' lack of attention and effort, both in class (e.g., "I work on other things when I'm in this class") and outside of class (e.g., "Outside of class, I don't put much work in on this course"). The item pool for *Emotional Disaffection* comprised 9 items depicting negative emotions about working on science, including boredom (e.g., "This class can be pretty dull") and worry (e.g., "This class is stressing me out"). Since each of the item pools contained items tapping multiple aspects of the construct (e.g., inside and outside of class), the key question was whether 5 items could be identified for each subscale that covered the relevant conceptual space while also showing a unidimensional structure and satisfactory internal consistencies.

Identity as a scientist. Building on previous research on academic identity and purpose in science (Saxton et al., 2014) and garden-based education (Skinner, Chi, & LEAG, 2012) from a self-determination perspective, three scales tapped students' science identity, or their deeply-held views of themselves and their potential to enjoy and succeed in science. The item pools for these subscales consisted of 14 items tapping (1) *Science Identity* or students' beliefs about their fit to science (e.g., "I am the kind of person who can succeed in science," "Sometimes I feel like I don't belong in science," reverse-coded); (2) students' future-oriented *Science Career Plans* ("I am planning on a job that involves science," "I'm just not cut out for a career in science," reverse-coded); and (3) a sense of *Purpose in Science* or the conviction that science makes important contributions to society (e.g., "Science can help solve many of society's problems"). Because these item pools contained both positively and negatively worded items, the primary question was whether 4-5 items could be identified for each subscale that were both unidimensional and internally consistent.

Positive relationships with peers. Finally, four newly developed items tapped students' perceptions about whether they had made supportive connections with peers in science class.

Final survey. The final survey, entitled *Self-determination, Purpose, Identity, and Engagement in Science* (or SPIRES) includes 11 short-form scales containing 4-5 items each (see Appendix). All survey items used a 5-point Likert scale, ranging from (1) *Not true at all* to (5) *Totally true*. Total scale scores for each construct were calculated by averaging items within a scale, with negatively-valenced items reverse-coded. All scales could range from 1 to 5 with higher scores indicating more of the respective construct (e.g., higher perceived competence or greater emotional disaffection).

Academic performance. Students' actual grades at the end of the class were provided by their instructors, and could range from 0 (grade of "F") to 4.0 (grade of "A").

Missing Data

Although few responses (1.187%) were missing at T1, Little's missing completely at random (MCAR) test (Little, 1988) indicated that those data were not MCAR (χ^2 (5580) = 6029.805, p > .001). Items near the end of the survey (focusing on engagement and disaffection) showed the highest percentage of missing data. Two separate Little's tests revealed that, although the engagement and disaffection items were not MCAR (χ^2 (1167) = 1343.824, p > .000), all remaining items were (χ^2 (1589) = 1641.570, p = .175). So, between T1 and T2, survey items were reorganized to place demographic items at the end. At T2, even fewer responses (.817%) were missing and Little's test for all items in the dataset indicated that the data were MCAR (χ^2 (3661) = 3499.363, p = .972). Hence, we concluded that missing values for engagement and disaffection at T1 were likely due to students dropping out before completing the survey. The expectation-maximization technique was used to impute missing data for both T1 and T2 and all analyses were conducted on the imputed datasets.

Data Analysis

CFA and measurement invariance analyses were performed in the R software environment using the lavaan package version 0.5-20 (Rossell, 2012). All factors were scaled through the use of a unit loading identification constraint and the reference variable was kept consistent across all models (CFA and measurement invariance discipline groups) for each scale. Models were evaluated for fit by considering the χ^2 test, the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), standardized root mean squared residual (SRMR), correlation residuals, and item loadings. The following cut points were used as general guidelines for interpreting model fit: CFI and TLI > .95 is considered good and > .90 is adequate; SRMR <.06 is good and < .08 is adequate (Hu & Bentler, 1999). Internal consistency analyses (Cronbach's alpha and McDonald's omega) were also performed in R. All remaining analyses used SPSS-version 24.

Results

Iterative Process of Item Selection

In order to construct short-forms of all 11 motivational subscales, an iterative confirmatory process was used to reduce each item pool from 6-12 per subscale to 4-5 per subscale. The items in each pool had already been adapted or generated to correspond to the set of pre-specified target constructs, so brief measures were constructed using conceptual and empirical criteria to identify the smallest subset of these conceptually-focused items (4-5 items) that showed the strongest psychometric properties. Items were initially selected based on three substantive goals: (1) to maximize construct validity by retaining the prototypical "anchor" items for each construct (e.g., "I am good at science" for *Competence*); (2) to retain the breadth of each construct's conceptual space by incorporating items that tapped all the sub-dimensions or valences of the construct covered by the original item pool (e.g., including items that marked both identified and intrinsic sub-dimensions of *Autonomy*); and (3) to maximize face validity by retaining items that had particular relevance to the context of undergraduate science courses (e.g., including "The material we cover in class is challenging (in a good way)" for *Emotional Engagement*).

After identifying candidate item bundles according to these conceptual criteria, a series of analyses were conducted to determine the psychometric properties of these substantively-comparable bundles. Cumulative evidence from these analyses was used to identify the final item set for each construct, which are presented below. First, single factor models were tested within the confirmatory factor analysis (CFA) framework to determine whether items composing each scale were "sufficiently homogenous" (McDonald, 1999, p. 175) to qualify as unidimensional. In contrast to previous measurement work in undergraduate science (e.g., Glynn et al., 2009), which typically uses EFAs to make sense of item pools that were not generated with specific *a priori* constructs in mind, CFAs were used in the current study, because all items had been generated or adapted to capture specific theoretically-derived constructs, that are consistent with the larger SDT literature. Moreover, the items parallel those of other instruments, commonly used outside the domain of science or with slightly younger students, as described previously, so that their dimensionality has been tested and is well understood. CFAs provide a more stringent test of the fit of theoretical models to the structure of item pools.

Second, internal consistencies (Cronbach's alpha and McDonald's omega) and cross-time stabilities (as lower-bound estimates of test-retest reliabilities) were calculated for each scale. Third, after T1 data were used to identify the strongest item bundles, the psychometrics of these scales were cross-validated with T2 data (McDonald, 1999). Fourth, scales were tested for measurement invariance across three science disciplines (biology, chemistry, and physics) at T1, by examining a series of nested models, each with more stringent constraints, testing for configural, metric, and scalar invariance. Fifth, scales were tested for evidence of predictive validity, by examining (1) inter-correlations among constructs; and (2) correlations with final letter grades. Correlational patterns were evaluated for congruence with expectations derived from SDT. Based on these correlations, a final check was made for discriminant validity: Any scales that showed correlations that approached the scales' levels of internal consistency were reviewed for item content overlap. After any such overlap was minimized, we considered the scales finalized.

Unidimensionality of Motivational Scales

In order to determine whether items composing each scale were unidimensional, single factor models were tested using CFA. Because of the large sample size (N_{T1} = 856; N_{T2} = 574) and the tendency for χ^2 to increase with sample size, we recognized that significant values for χ^2 could reflect small model-data discrepancies, and therefore, might not indicate differences of practical or theoretical significance. However, because researchers have been criticized for dismissing the χ^2 test on grounds of sample size (Kline, 2011), we considered this information in the context of other indicators of fit. Moreover, although we report the RMSEA, we did not find this particular statistic useful for making determinations regarding model fit for these brief scales, because multiple simulation studies have found the RMSEA to be sensitive to model size-- with models comprised of fewer items and factors, like our 4-5 item one-factor models, at a particular disadvantage (Breivik & Olsson, 2001; Cheung & Rensvold, 2002).

Results of the CFA analyses are presented in Supplementary Table S1 and S2. Four scales showed cumulative evidence of good fit at both time points: Emotional Engagement, Science Career Plans[1], Purpose, and Positive Relationships with Peers. These scales, despite their significant χ^2, all had fit indices (CFI, TLI, SRMR) that fell within conventional guidelines for good fit, had factors loadings that were all greater than .35 (Tabachnick & Fidell, 2000), and correlation residuals that were all less than .10. Five other scales showed cumulative evidence of good to adequate fit at both time points: Autonomy, Behavioral Engagement, Behavioral Disaffection, Emotional Disaffection, and Science Identity. These scales (all of which incorporated items from multiple sub-dimensions or valences) generally had CFI and SRMR fit indices that fell within the good range, TLI values that fell within the adequate range, generally no loadings less than .35, and no pattern of correlation residuals above .10 (i.e., no pattern consisting of 2 or more at a single time point). The Behavioral Disaffection scale did have one item loading at T1 that fell below .35, however, at T2 all loadings were above .35.

Of the 11 scales, two fell below the adequate range on some indicators of unidimensionality on at least one time point: Competence (which included items about the class

[1] The Science Career Plans scale, as seen in Table S2, did have one correlation residual (-.101) above the .10 cut mark; however, given that this value is so close to the cut mark and that all fit indices indicate good fit, we determined that this scale was best categorized as having good fit at both time points.

and about science in general) and Relatedness (which included items about fitting in with students in the class, as well as in the class itself and the major). Competence had good to adequate fit at T1, but had a TLI at T2 that fell below the adequate range, at least one item loading that fell below .35 at both time points, and there were two correlation residuals above .10 at T2. In the case of Relatedness, the CFI had good to adequate fit at both time points and the item loadings were all above .35, but the TLI missed adequate by .003 or less at both time points and two correlation residuals were greater than .10 at T2. Hence, we concluded that 9 of the 11 scales showed adequate to good cumulative evidence for the unidimensionality of their item sets at both time points, whereas two scales showed adequate evidence of unidimensionality at T1, but not in the cross-validation at T2.

Reliability of Motivational Scales

In a second step, we examined multiple indicators of reliability: two kinds of internal consistency (Cronbach's alpha and McDonald's Omega) and cross-time stability (as a lower-bound estimate of test-retest reliability). As shown in Table 4, these statistics were generally satisfactory: Cronbach's alphas ranged from .67 to .91 at T1, and from .73 to .94 at T2; McDonald's Omegas, which were higher than the alphas for all scales and time points, ranged from .74 to .94 at T1, and from .82 to .95 at T2; cross-time stabilities ranged from .49 to .79. Six scales showed strong evidence of internal consistency, with alphas and omegas at both time points that were at or above .80. Four other scales showed satisfactory internal consistencies, with alphas greater than .70 and omegas at or above .80. One scale, namely, Behavioral Disaffection, had mixed results: It showed an alpha at T1 that fell below satisfactory levels (α =.67); however, the alpha at T2 and the omega at both time points exceeded .70. For all scales, the cross-time stabilities were consistently high, indicating that test-retest reliabilities were satisfactory.

Insert Table 4 about here

Measurement Invariance across Disciplines

In the third step of the analyses, we examined measurement invariance across disciplines (biology, physics, and chemistry) for all scales at T1, by testing a series of three nested models, each with more stringent constraints, examining configural, metric, and scalar invariance. Based on evidence from these models (see Tables S3-5), each scale could be classified as "unambiguously invariant," "strong evidence for invariance," or "moderate evidence for invariance." Three scales, Behavioral Engagement, Science Identity, and Science Career Plans, met criteria for "unambiguously invariant:" The $\Delta\ \chi^2$ was non-significant across progressively constrained models, ΔCFI $\leq$ -.01 (Cheung & Rensvold, 2002), and other fit indices (excluding RMSEA) either remained in or moved into levels generally considered indicative of good fit.

Six scales meet criteria for "strong evidence of invariance:" Autonomy, Relatedness, Emotional Engagement, Emotional Disaffection, Purpose in Science, and Positive Relationships with Peers. These scales all had at least one significant chi-square difference test (which, as noted, are sensitive to a large sample size) and some had a $\Delta CFI_{\text{metric to scalar}}$ that exceeded the -.01 cutoff. These scales also had fit that varied slightly as the models were progressively constrained, but nevertheless remained in the same cutoff designation (good or adequate) as the original CFA. Finally, two scales, namely, Competence and Behavioral Disaffection, showed "moderate evidence for invariance," in that, at some point in the progressive models, fit dropped from good to adequate, although overall fit of the constrained models still met adequate criteria.

Predictive Validity: Inter-correlations among Motivational Scales and Academic Performance

Means, standard deviations, skewness, and kurtosis for each scale appear in Table 5. There is no evidence for ceiling or floor effects on most of the scales at either time point. However, the Purpose subscale seems to be somewhat leptokurtic (kurtosis > 3) and negatively skewed, indicating that more students viewed science as very high in its contributions to solving societal problems. Given that the majority of students in this sample were science majors, this finding is not surprising. Table 6 presents predictive validity information, namely, the correlations among the motivational scales and between motivational scales and final course grades. As can be seen, correlations among the motivational scales at both time points showed exactly the pattern predicted by SDT (see Figure 1). All three self-systems were positively correlated with both Behavioral and Emotional Engagement (r = .408, averaged across components, self-systems, and time points), and negatively correlated with both Behavioral and Emotional Disaffection (average r = -.395). This pattern of connections indicated that, as posited by SDT, students who felt more competent, autonomous, and related also reported the kinds of engagement that are typical of a “motivated” student, namely, higher levels of behavioral and emotional engagement; it also indicated that students who felt that their needs were not met reported higher levels of both kinds of disaffection.

Insert Tables 5 and 6 about here

In the same vein, the components of Identity as a Scientist were positively inter-correlated with each other: Science Identity with Career Plans (r = .569, averaged across both time points) and Purpose (average r = .382) as well as Purpose and Career Plans (average r = .240). And, consistent with SDT, all three self-system appraisals were correlated positively and significantly with Science Identity (rs = .699, .481, and .722, for Competence, Autonomy, and Relatedness, respectively, averaged across time points), Science Career Plans (rs = .502, .375, and .338, for Competence, Autonomy, and Relatedness, respectively), and Sense of Purpose (rs = .271, .302, and .320, for Competence, Autonomy, and Relatedness, respectively) indicating, for example, that students who felt a stronger calling to science (Science Identity) and whose needs were met in science coursework were more likely to report that they planned to integrate science into their future careers. And Science Identity, Science Career Plans, Purpose in Science, and supportive Peer Relationships were also, in turn, correlated with components of engagement, especially Emotional Engagement (r = .275, averaged across constructs and time points).

Correlations with academic performance. As expected, the motivational scales also showed evidence of predictive validity with academic performance, in that scores on all scales at T2 were correlated significantly with final course grades. Most scales at T1 were also correlated with final grades, although not surprisingly, connections were consistently stronger with measures from T2 near the end of the term (rs ranged from -.279 - .400) than from T1 at the start of term (rs ranged from -.184 - .196). All three self-system processes were positive correlates of grades at both time points: Competence (r = .297, averaged across time points), Autonomy (average r = .203), and Relatedness (average r = .246); as were both components of engagement (average r = .206, averaged across behavior and emotion) and both components of disaffection (average r = -.204) at both time points, consistent with the notion that engagement may contribute to higher levels of learning and performance. Final grades were also significantly correlated with two of the three components of Identity as a Scientist at both time points: Science

Identity (r = .237 averaged across time points) and Purpose in Science (average r = .119); Science Career Plans and Peer Relationships at T1 were not correlated significantly with final grades, but both were correlated positively and significantly with grades by T2 (both rs = .128, $p < .05$).

Discussion

As articulated by expert panels, an important first step in improving undergraduate performance and persistence in STEM is identifying the factors that contribute to these outcomes-- factors that research increasingly suggests are not only cognitive and pedagogical, but also motivational and interpersonal. Because a necessary part of this process is developing tools that can capture these factors empirically, the goal of the current study was to create a set of brief conceptually-focused and psychometrically-sound measures of key motivational processes in undergraduate science coursework. These scales complement existing measures that focus more on social cognitive factors or on discipline-specific attitudes and skills, by broadening our view of the student motivational experience—to consider questions not only about confidence and efficacy, but also about personal autonomy and commitment, belongingness, engagement, identity, and sense of purpose.

Performance of the brief measures of self-appraisals, engagement, and identity. These 11 scales generally showed strong evidence of reliability and predictive validity. They were internally consistent (as shown by satisfactory levels of Cronbach's alpha and McDonald's omega), homogenous (as indicated by CFA tests of unidimensionality), and stable across time (indicating high test-retest reliability). Moreover, all scales correlated with each other as predicted by SDT: Students with higher self-appraisals (of competence, relatedness, and autonomy) and identity as scientists (science identity, purpose, and career plans) also reported higher levels of behavioral and emotional engagement, and lower levels of behavioral and emotional disaffection. And all these motivational factors correlated with students' actual grades in their courses at Time 2; as did all but two (career plans and peer relationships) measured at least seven weeks earlier at Time 1. Moreover, tests of measurement invariance suggest that these scales can be used for students from biology, chemistry, and physics classes. Measures that span science disciplines can contribute to efforts to compare and contrast discipline-based instructional strategies, thus creating a broader evidence base from which educators can draw.

At the same time, of the 11 scales, three showed less than ideal measurement properties on at least one of the specific indicators of unidimensionality, reliability, or invariance at one of the time points. The item set comprising Competence showed less than adequate unidimensionality on one indicator (TLI) at T2, and both Competence and Relatedness showed two correlation residuals above cut-offs at T2. For Behavioral Disaffection, the loading of one item fell below conventional cut-offs at T1, and one indicator of internal consistency (Cronbach's alpha) fell below conventional cut-offs at T1. And Competence and Behavioral Disaffection showed only moderate evidence for invariance across disciplines. When considering how this might affect the functioning of these three scales, we looked across the other psychometric indicators and time points. All three scales showed adequate performance on the specific indicators at the other time point, and on other indicators of unidimensionality and reliability at both time points. For example, Competence showed good unidimensionality according to the other indicator (CFI) at T2, and adequate or good unidimensionality according to the both indicators at T1, and both Competence and Relatedness showed satisfactory alphas and omegas at both time points (indicating satisfactory homogeneity within item sets), and high cross-time stability. All items tapping Behavioral disaffection showed adequate loadings at T2;

satisfactory internal consistency according to the other indicator (McDonald's omega) at T1 and both indicators at T2; and high cross-time stabilities. And the overall fit of the constrained invariance models for Competence and Behavioral Disaffection still met conventional criteria. Most importantly, even if the unidimensionality and reliability of these brief scales could only be considered adequate, this did not seem to interfere with their predictive validity: All three were correlated robustly with other markers of students' motivation, and Competence and Relatedness were the strongest predictors of students' actual course grades at T2. Hence, we tentatively concluded that, despite their imperfections, all 11 scales could be considered ready for further use and testing.

Sense of Relatedness and Science Identity. Finally, two of the 11 scales, namely, Relatedness and Science Identity, showed signs that discriminant validity may be low. Even though each showed good evidence of reliability and predictive validity on its own, the two were highly correlated (.740 and .703 at T1 and T2, respectively), indicating considerable overlap. Conceptually, the two constructs are distinguishable—Relatedness refers to a sense of belonging and inclusion with classmates, and in science classes and the science major (generally indicating a feeling that "This is my tribe"), whereas Science Identity refers to students' deeply-held views of themselves and their potential to connect with and succeed in the community of science (generally indicating that "This is who I am"). Hence, this empirical overlap can be interpreted in at least three different (but not mutually exclusive) ways. It could indicate that students do not distinguish between these two constructs, in which case item sets could be combined. Or, it could indicate that further empirical differentiation is needed, perhaps within Relatedness—such that separate Relatedness subscales could focus on a sense of inclusion with *classmates* (e.g., "I fit in well with the other students in this class"), fit with *science classes* (e.g., "This course is a good place for students like me"), and a sense of belonging in the *major* (e.g. "I'm not really sure that science is the right major is for me") Although, as shown in this study, these items can be combined to form a relatively unidimensional, homogeneous, and well-functioning scale, separate subscales might be even more unidimensional and homogeneous, and show differential connections to Science Identity.

Alternatively, a third possibility is that the high correlations between Relatedness and Science Identity could indicate that the most important contributor to the development of students' identities as scientists is their sense of belongingness, welcome, and inclusion with their classmates and in the classroom and major. Consistent with this interpretation, Relatedness is also the strongest correlate at both time points of the two other aspects of Identity as a Scientist, namely, Science Career Plans and Purpose in Science, even though neither of them share any conceptual or item content overlap with Relatedness. Moreover, other self-appraisals besides Relatedness also show strong connections to Science Identity, most noticeably Competence (with correlations of .695 and .703 at T1 and T2, respectively), although Competence shares no conceptual or item content overlap with Science Identity either. Hence, we tentatively conclude that educators, interventionists, and researchers should consider the role of Relatedness, a construct that is highlighted by SDT but missing from social-cognitive and expectancy-value models, in shaping the development of a crucial motivational resource, namely, a strong science identity. A positive science identity not only seems to promote engagement and protect students from disaffection during a specific science class, but it may also strengthen persistence and resilience in the face of subsequent obstacles and challenges (Chang et al., 2011). In this case, future research should explicitly investigate the kinds of learning activities, pedagogical strategies, and interpersonal relationships that support the development of a sense of relatedness and belonging in science, especially for students who are first generation, women, and from ethnic and racial minority groups and low socioeconomic backgrounds, who currently may not feel welcome in science courses, majors, or careers (Carlone & Johnson, 2007). One clue in this regard, at least in the current study, may be found in the significant correlation between a sense of Relatedness, on the one hand, and students' positive relationships with classmates, on the other (r = .344 at T1 and .341 at T2).

Study Strengths and Limitations

The current study had both strengths and limitations. Although a notable strength was the reliance on a strong theory of motivation (based on a robust body of evidence), neither the theory nor the measures are comprehensive. For example, the suite of measures referred to only one facet of the undergraduate experience, namely, science classes. It did not tap students' experiences in other program-related activities, such as undergraduate research, advising, outreach, or clubs. It is possible that specific experiences, such as working in a faculty research lab with other students, would be *more* likely to fulfill students' motivational needs than participation in the typical science class (Eagan et al., 2013). In terms of the sample, a strength of the present study was its inclusion of 8 classes representing 3 disciplines, and the relative diversity of the sample across some demographic groups. At the same time, however, all of the students were drawn from one institution, thus potentially limiting the generalizability of the findings. Finally, to expand on efforts to validate these (and other) motivational scales, additional data sources are needed in future studies, such as qualitative student interviews, classroom observations, information from instructors, or additional markers of student motivation (e.g., attendance in tutorials, completion of extra-credit assignments).

Educational Implications and Future Research

A primary goal of the current study was to create a set of measures that were high in face validity and brief enough for instructors to use repeatedly in their classes. The current scales are an important step in this direction: The items are written in plain language that makes their meaning clear and students can complete them all in about 10 minutes. Together, these motivational scales paint a picture for instructors of the undergraduate experience that, based on the assumptions of SDT, highlights the questions students may be asking themselves in their STEM classes: “Do I have what it takes to succeed in science?” (competence); “Am I personally committed to the hard work a science major entails?” (autonomy); “Do I belong here?” (relatedness); “Am I the kind of person who is a good fit with this discipline and profession?” (identity); and “Is the work of science relevant and worthwhile?” (purpose). These are questions that instructors themselves likely answered in the affirmative many years ago, but they may have done so implicitly without really reflecting on the issues such questions entail. The utilization of these scales as a regular part of their teaching provides instructors the opportunity to consider the concerns that preoccupy many of their students. Perusing students’ responses may also encourage instructors to think more carefully about the role that they themselves play in helping students wrestle with these questions—through the academic activities and supports they provide. Such reflections may also awaken instructor interest in professional development (PD) activities designed to improve the student experience (Borrego, & Henderson, 2014).

Educational interventions. A second goal of the current study was to construct measures that would guide interventionists in creating and testing interventions designed to improve undergraduate STEM education. The suite of motivational measures may help interventionists calibrate or select among alternative pedagogical strategies, ensuring that the practices they promulgate are effective in meeting student needs and bolstering student engagement. SDT may also be useful in providing a lens to focus the variety of PD activities that are available to STEM instructors. It suggests, for example, that many of the pedagogical strategies identified by discipline-based educational research, such as those emphasizing active and cooperative learning, may exert their impact on performance at least partly through their effects on students' self-systems, their engagement, or supportive peer relationships. Instructors may come to appreciate that the combination of authentic academic work and supportive instructor practices seems to be especially important to students from underrepresented minority groups and first-generation college students (Wigfield et al., 2015).

Future research. A third goal of the current study was to construct measures that would be useful to educational researchers in creating process-oriented accounts of students' experiences in STEM classes. Such studies could eventually link the features of teaching that are under the institution's control (e.g., pedagogical strategies, learning activities, and instructor supports) to the outcomes of value to those institutions, such as deep learning and persistence in STEM majors. The measures developed here would allow researchers to finish answering important questions, such as whether specific self-systems or features of identity play a bigger role in the engagement or persistence of students from particular groups. One interesting hypothesis would be the notion that a sense of relatedness and identity are particularly important to groups of students who have historically been marginalized in STEM coursework and professions.

Pedagogical practices and interpersonal relationships. The larger theoretical framework provided by SDT also creates a bridge to the next logical educational, intervention, and research steps in this work, namely, to more intentionally improve science teaching and learning by identifying the pedagogical and interpersonal factors that promote students' self-system appraisals, engagement, and identities as scientists. According to SDT, there are two primary levers that can be used to transform students' motivational experiences in STEM classes (see Table 1). The first is the nature of the academic work that students are required to undertake. When academic work is active and authentic, that is, hands-on, heads-on, experiential, project-based, relevant, progressive, and integrated across subject matter, it becomes intrinsically motivating, inherently interesting, and engaging. Such learning activities can be demanding, but they also help students rise to the challenge (Stefanou, Stolk, Prince, Chen, & Lord, 2013). This kind of active and authentic academic work has been studied for decades in research on motivation (Darling-Hammond et al., 2008), and they are the same kinds of pedagogical approaches that have been shown to be successful in improving engagement and achievement in undergraduate STEM coursework (Singer & Smith, 2013).

A second lever, also supported by a robust body of motivational research, highlights the quality of the interpersonal relationships that students develop with their teachers and peers (Martin & Dowson, 2009; Wentzel, 2009; Wentzel & Muenks, 2016). According to SDT (and many other motivational theories), supportive relationships with teachers and peers are the basis from which students develop positive self-system processes, a strong academic identity, and motivational resilience. From this perspective, teachers are supportive to the extent they foster caring relationships, provide challenging learning activities with high expectations and clear feedback, and explain the relevance of activities and rules while soliciting input from students and respecting their opinions. Peers are supportive to the extent they include, connect with, listen to, and work constructively with others, both inside and outside the classroom. Discipline-based educational research in STEM also highlights these relationships as central to students' STEM experiences (Kuh, 2007). Taken together, decades of research at a variety of educational levels (as well as the current study) suggest that instructors, interventionists, and researchers may find it fruitful to consider students' motivation, and the supports that promote its optimization, in their efforts to improve the quality of undergraduates' learning experiences and success in STEM.

References

Adedokun, O. A., Bessenbacher, A. B., Parker, L. C., Kirkham, L. L., & Burgess, W. D. (2013). Research skills and STEM undergraduate research students' aspirations for research careers: Mediating effects of research self□efficacy. *Journal of Research in Science Teaching*, *50*(8), 940-951.

American Society for Engineering Education (ASEE). (2009). *Creating a culture for scholarly and systematic innovation in engineering education: Ensuring U.S. engineering has the right people with the right talent for a global society.* Washington, DC: Author.

American Society for Engineering Education (ASEE). (2012). *Innovation with impact: Creating a culture for scholarly and systematic innovation in engineering education*. Washington, DC: Author.

Bauer, C. F. (2005). Beyond" student attitudes": Chemistry Self-concept Inventory for assessment of the affective component of student learning. *Journal of Chemistry Education*, *82*(12), 1864.

Borrego, M., & Henderson, C. (2014). Increasing the use of evidence□based teaching in STEM higher education: A comparison of eight change strategies. *Journal of Engineering Education*, *103*(2), 220-252.

Breivik, E. & Olsson, U.H. (2001). Adding variables to improve fit: the effect of model size on fit assessment in LISREL. In R. Cudeck, S. du Toit S , D. Sörbom, (Eds.), *Structural Equation Modeling: Present and Future* (pp. 169-194). Lincolnwood, IL: Scientific Software International.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science*

Teaching, *44*(8), 1187-1218.

Chang, M. J., Eagan, M. K., Lin, M. H., & Hurtado, S. (2011). Considering the impact of racial stigmas and science identity: Persistence among biomedical and behavioral science aspirants. *The Journal of Higher Education*, *82*(5), 564.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255.

Chi, U. (2013). *Classroom engagement as a proximal lever for student success in higher education: What a self-determination framework within a multi-level developmental system tells us*. Unpublished doctoral dissertation, Portland State University.

Christenson, S. L., Reschly, A. L., & Wylie, C. (2012). *Handbook of research on student engagement.* Springer: New York, NY.

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy and relatedness: A motivational analysis of self-system processes. In M. Gunnar & L. A. Sroufe (Eds.), *Minnesota Symposium on Child Psychology: Vol. 23. Self processes in development* (pp. 43-77). Chicago: University of Chicago Press.

Dalgety, J., Coll, R. K., & Jones, A. (2003). Development of Chemistry Attitudes and Experiences Questionnaire (CAEQ). *Journal of Research in Science Teaching*, *40*(7), 649-668.

Darling-Hammond, L., Barron, B., Pearson, P. D., Schoenfeld, A. H., Stage, E. K., Zimmerman, T. D., Cervetti, G. N., & Tilson, J. L. (2008). *Powerful learning: What we know about teaching for understanding*. San Francisco, CA: Jossey-Bass.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.

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.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Thousand Oaks, CA: Sage.

Eagan, M. K., Hurtado, S., Chang, M. J., Garcia, G. A., Herrera, F. A., & Garibay, J. C. (2013). Making a difference in science education: The impact of undergraduate research programs. *American Educational Research Journal*, *50*(4), 683-713.

Espinosa, L. (2011). Pipelines and pathways: Women of color in undergraduate STEM majors and the college experiences that contribute to persistence. *Harvard Educational Review*, *81*(2), 209-241.

Gasiewski, J. A., Eagan, M. K., Garcia, G. A., Hurtado, S., & Chang, M. J. (2012). From gatekeeping to engagement: A multicontextual, mixed method study of student academic engagement in introductory STEM courses. *Research in Higher Education*, *53*(2), 229-261.

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science Motivation Questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching*, *48*(10), 1159-1176.

Glynn, S. M., & Koballa, T. R. (2006). Motivation to learn in college science. In J. Mintzes (Ed.) *Handbook of College Science Teaching* (pp. 25-32). Arlington, VA: National Science Teachers Association Press.

Glynn, S. M., Taasoobshirazi, G. & Brickman, P. (2009). Science Motivation Questionnaire: Construct validation with nonscience majors. *Journal of Research in Science Teaching, 46*, 127-146.

Handelsman, M. M., Briggs, W. L., Sullivan, N., & Towler, A. (2005). A measure of college student course engagement. *The Journal of Educational Research*, *98*(3), 184-192.

Hawwash, K. (2007). Attractiveness of education. In C. Borri & F. Maffioli (Eds.), TREE: Teaching and research in engineering in Europe: Re-engineering engineering education in Europe. Florence, Italy: Firenze University Press.

Hu, L. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Hurtado, S., Newman, C. B., Tran, M. C., & Chang, M. J. (2010). Improving the rate of success for underrepresented racial minorities in STEM fields: Insights from a national project. *New Directions for Institutional Research*, *2010*(148), 5-15.

King, R. (2008). Addressing the supply and quality of engineering graduates for the new century: Surry Hills, NSW, Australia: Carrick Institute.

Kline, R.B. (2011). *Principles and practice of Structural Equation Modeling*. New York, NY: The Guilford Press.

Kuh, G. D. (2007). What student engagement data tell us about college readiness. *Peer Review*, *9*(1), 4.

Lee, W., & Reeve, J. (2012). Teachers' estimates of their students' motivation and engagement: Being in synch with students. *Educational Psychology*, *32*(6), 727-747.

Lee, J. K., Alston, A. T., & Kahn, K. B. (2015). Identity threat in the classroom: Review of women's motivational experiences in the sciences. *Translational Issues in Psychological Science*, *1*(4), 321-330.

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.

Liu, O. L., Bridgeman, B., & Adler, R. M. (2012). Measuring learning outcomes in higher education: Motivation matters. *Educational Researcher*, *41*(9), 352-362.

Martin, A. J., & Dowson, M. (2009). Interpersonal relationships, motivation, engagement and achievement: Yields for theory, current issues, and educational practice. *Review of Educational Research, 79*, 327-365.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

National Academy of Engineering (NAE). (2004). The engineer of 2020. Washington, DC: National Academies Press.

Olson, S., & Riordan, D. G. (2012). Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Washington, DC: President's Council of Advisors on Science and Technology.

President's Council of Advisors on Science and Technology. (2012). Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Washington, DC: Author.

Reeve, J. (2012). A self-determination theory perspective on student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 149-172). New York: Springer.

Robbins, S. B., Lauver, K., Le, H. David, D., Langley, R., & Carlstom, A. (2004). Do psychosocial and study skills factors predict college outcomes? A metaanalysis. *Psychological Bulletin, 130*(2), 261-288.

Rossell, Y. (2012) lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36.

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57*, 749-761.

Ryan, R. M., & Deci, E. L. (2016). Facilitating and hindering motivation, learning, and well-being in schools: Research and observations from Self-Determination Theory. *Handbook of Motivation at School*, 2nd ed (pp. 96-119). Malwah, NJ: Erlbaum.

Ryan, R. M., & Deci, E. L. (2017). Self-determination theory: Basic psychological needs in motivation, development, and wellness. *New York: Guilford Publishing*.

Saxton, E., Burns, R., Holveck, S., Kelley, S., Prince, D., Rigelman, N., & Skinner, E. A. (2014). A Common Measurement System for K-12 STEM education: Adopting an educational evaluation methodology that elevates theoretical foundations and systems thinking. *Studies in Educational Evaluation*, *40*, 18-35.

Schunk, D. H., & DiBenedetto, M. K. (2016). Self-efficacy theory in education. In K. Wentzel & D. Miele (Eds.), *Handbook of Motivation at School*, 2nd ed (pp. 34-54). Malwah, NJ: Erlbaum.

Singer, S., & Smith, K. A. (2013). Discipline□based education research: Understanding and improving learning in undergraduate science and engineering. *Journal of Engineering Education*, *102*(4), 468-471.

Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (Eds.). (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, DC: National Academies Press.

Skinner, E. A., Chi, U., & the Learning-Gardens Educational Assessment Group (2012). Intrinsic motivation and engagement as "active ingredients" in garden-based education:

Examining models and measures derived from self-determination theory. *Journal of Environmental Education, 43(1),*16-36.

Skinner, E. A., Kindermann, T. A., & Furrer, C. (2009a). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69,* 493-525.

Skinner, E. A., Kindermann, T. A., Connell, J. P., & Wellborn, J. G. (2009b). Engagement as an organizational construct in the dynamics of motivational development. In K. Wentzel & A. Wigfield (Eds.), *Handbook of Motivation at School* (pp. 223-245). New York, NY: Routledge.

Skinner, E. A., Wellborn, J. G., & Connell, J. P. (1990). What it takes to do well in school and whether I've got it: The role of perceived control in children's engagement and school achievement. *Journal of Educational Psychology, 82,* 22-32.

Stefanou, C., Stolk, J. D., Prince, M., Chen, J. C., & Lord, S. M. (2013). Self-regulation and autonomy in problem-and project-based learning environments. *Active Learning in Higher Education*, *14*(2), 109-122.

Tabachnick, B. G., & Fidell, L. (2000). Using multivariate statistics (4th edn.). Boston: Allyn & Bacon.

Tsui, L. (2007). Effective strategies to increase diversity in STEM fields: A review of the research literature. *The Journal of Negro Education*, 555-581.

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., & Vallieres, E. F. (1992). The Academic Motivation Scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement*, *52*, 1003-1017.

Wentzel, K. R. (2009). Students' relationships with teachers as motivation contexts. In K. Wentzel & A. Wigfield (Eds.), *Handbook of motivation in school* (pp. 301-322). Malwah, NJ: Erlbaum.

Wentzel, K. R., & Muenks, K. (2016). Peer influence on students' motivation, academic achievement, and social behavior. In K. R. Wentzel & G. B. Ramani (Eds.), *Handbook of social influences in school contexts: Social-emotional, motivation, and cognitive outcomes* (pp. 13-30). New York, NY: Routledge.

Wigfield, A., Eccles, J. S., Fredricks, J. A., Simpkins, S., Roeser, R., & Schiefele, U. (2015). Development of achievement motivation and engagement. In R. M. Lerner (Series Ed.) & M. Lamb (Volume Ed.), *Handbook of child psychology and developmental science, 7th Ed. Vol.3. Socioemotional processes* (pp. 657-700). New York: John Wiley.

Wigfield, A., Tonks, S., & Klauda, S. L. (2016). Expectancy-value theory. In K. R. Wentzel & A. Wigfield (Eds). *Handbook of motivation at school* (pp. 55-75). New York: Routledge.

Zusho, A., Pintrich, P. R., & Coppola, B. (2003). Skill and will: The role of motivation and cognition in the learning of college chemistry. *International Journal of Science Education*, *25*(9), 1081-1094.

Table 1.

A Self-Determination Theory Perspective on the Student Experience in STEM Classes

Students' Views of Themselves (Self-System Processes)

- *Competence/Efficacy*: Students' beliefs about whether they have the ability to succeed in STEM classes and fields.
- *Autonomy/ Ownership*: Whether students are personally committed to the work in STEM classes and careers.
- *Relatedness/Belonging*: Whether students feel that "people like them" are welcome and would be accepted in the study and professions of STEM.

Student Engagement in STEM Academic Work

- *Academic engagement*: Whether students show high quality participation in learning activities, including effort (hard work, exertion, follow-through) and enthusiasm (interest, curiosity).

Student Identity as a Scientist

- *Science identity*: Students' deeply-held views of themselves and their potential to enjoy and succeed in STEM classes and careers.
- *Purpose in Science*: Whether students are convinced that classwork and professional work in STEM is meaningful, important, and worthwhile.
- *Science Career Plans*: Whether students view STEM as a key part of their future vocational plans.

Levers of change: Authentic academic work

Rationale: Active participation, engagement, and effort are promoted by interesting authentic tasks that matter to the larger community.

Components: Academic work is authentic to the extent it is hands-on, heads-on, experiential, project-based, authentic, relevant, progressive, and integrated across subject matter, or in other words, intrinsically motivating, inherently interesting, and fun.

Levers of change: Supportive relationships with teachers and peers

Rationale: Supportive relationships with teachers and peers are the basis upon which students construct a positive academic identity and develop motivational resilience.

Components: Teachers are supportive to the extent they foster caring relationships, provide challenging learning activities with high expectations and clear feedback, and explain the relevance of activities and rules while soliciting input from students and respecting their opinions. Peers are supportive to the extent they include, connect with, listen to, and work constructively with others, both inside and outside the classroom.

Table 2.

Student-reported Gender Composition and Ethnic Background

	Time Point	
	T1	**T2**
Gender		
Female	59%	61%
Male	41%	38%
Other	< 1%	< 1%
Ethnicity		
African American	4%	5%
Asian/Pacific Islander	18%	19%
Caucasian	62%	60%
Hispanic/Latino	11%	12%
Native American/Native Alaskan	2%	2%
Rather not say	3%	5%

Note. The ethnicity percentages do not total to 100% because students were invited to '*mark all that apply'* on the surveys.

Table 3.

Measures based on Self-determination Theory-based from which the Subscales of the Self-determination, Purpose, Identity, and Engagement in Science (SPIRES) were Adapted

Construct		*Measure*	*Example items*
Self-systems Processes			
1.	*Competence*	SPIRES	Even if they are challenging, I can do well in my science classes. I don't have the intelligence/brains to succeed in science. (-)
	Perceived Control	Skinner, Wellborn, & Connell, 1990	I can do well in school if I want to. I don't have the brains to do well at school. (-)
2.	*Autonomy* • *Identified* • *Intrinsic* • *Amotivation*	SPIRES	Why do I do my classwork and homework for this course? • Because doing well in science is important to me. • Because it's fun to answer challenging science questions. • For this class, I just learn the stuff I have to in order to pass the test(s).
	Autonomy • *Identified* • *Intrinsic* • *Amotivation*	Ryan & Connell, 1989 Vallerand et al., 1992	Why do I do my classwork? • Because it's important to. • Because it's fun. • I honestly don't know; I really feel like I am wasting my time in college.
3.	*Relatedness*	SPIRES	I fit in well with the other students in this class. In science courses, I feel like an outsider. (-)
	Relatedness	Chi, 2013	I can relate to the other students in this class. In this class, I feel like an outsider. (-)
Engagement vs. Disaffection			
4.	*Behavioral Engagement*	SPIRES	I try hard to do well in this class.

		Skinner et al., 2009	I try hard to do well in school.
5.	*Emotional Engagement*	SPIRES	I enjoy the time I spend in this class.
		Skinner et al., 2009	I enjoy learning new things in class.
6.	*Behavioral Disaffection*	SPIRES	I work on other things when I'm in this class.
		Skinner et al., 2009	When I'm in class, I think about other things.
7.	*Emotional Disaffection*	SPIRES	When in class, I feel bored.
		Skinner et al., 2009	When we work on something in class, I feel bored.
Identity as a Scientist			
8.	*Science identity*	SPIRES	I am the kind of person who can succeed in science. Sometimes I feel like I don't belong in science. (-)
	Identity	Saxton et al., 2014	I am the kind of person who can succeed in Math/Science. Math/Science doesn't have anything to do with me. (-)
9.	*Science Career Plans*	SPIRES	I am planning on a job that involves science.
	Identity	Saxton et al., 2014	I want to be a scientist/ mathematician when I grow up.
10.	*Purpose in Science*	SPIRES	I believe that science can help make the world a better place.
	Sense of Purpose	Skinner, Chi, & LEAG, 2012	By gardening, we can make the world a better place.
11.	*Positive Peer Relationships and Collaboration*	SPIRES	I have gotten to know other students in this class.
	Peer Relationships	Skinner, Chi, & LEAG, 2012	I feel comfortable with the kids at school.

Table 4.

Internal Consistencies (Cronbach's alpha (α) and McDonald's omega (ω)) at both T1 and T2 and Cross-time Stabilities from T1 to T2 for All Scales

Construct	No. of Items	Internal Consistency Reliability				Cross-Time Stability T1 to T2
		Fall T1		Fall T2		
		α	ω	α	ω	
Self-System Processes						
Competence	5	.73	.82	.74	.82	.65***
Autonomy	5	.80	.85	.83	.91	.58***
Relatedness	5	.71	.83	.73	.82	.62***
Engagement vs. Disaffection						
Behavioral Engagement	5	.76	.80	.78	.84	.49***
Emotional Engagement	5	.82	.86	.84	.87	.64***
Behavioral Disaffection	5	.67	.74	.76	.83	.63***
Emotional Disaffection	5	.75	.81	.82	.86	.64***
Identity as a Scientist						
Science Identity	5	.80	.85	.83	.87	.79***
Science Career Plans		.80	.86	.82	.88	.69***
Purpose in Science	4	.91	.94	.94	.95	.53***
Positive Relationship with Peers	4	.87	.90	.90	.93	.61***

Note. Cross-time stability $N = 417$.

*** $p < .001$.

Table 5.

Means and Standard Deviations for All Survey Scales at Both Time Points.

Construct	Time point 1					Time point 2				
	M	*SD*	*Range*	*Skewness*	*Kurtosis*	*M*	*SD*	*Range*	*Skewness*	*Kurtosis*
Self-System Processes										
Competence	3.929	.712	3.60	-.623	.171	3.786	.766	3.80	-.436	-.309
Autonomy	4.041	.768	4.00	-.831	.424	3.824	.894	4.00	-.783	-.230
Relatedness	3.878	.747	4.00	-.537	-.241	3.706	.846	4.00	-.464	-.364
Engagement vs. Disengagement										
Behavioral Engagement	4.316	.630	3.20	-1.007	.626	4.093	.760	4.00	-.906	.584
Emotional Engagement	3.726	.816	4.00	-.432	-.226	3.456	.927	4.00	-.416	-.240
Behavioral Disengagement	1.592	.611	3.40	1.303	1.756	1.955	.842	4.00	1.149	1.387
Emotional Disengagement	2.233	.797	4.00	.819	.424	2.684	1.008	4.00	.517	-.435
Identity as a Scientist										
Science Identity	4.097	.775	3.60	-.879	.349	4.035	.847	4.00	-.964	.622
Science Career Plans	4.228	.925	4.00	-1.496	1.733	4.245	.945	4.00	-1.522	1.632
Purpose in Science	4.561	.730	4.00	-1.978	3.866	4.564	.756	4.00	-2.064	4.441
Positive Relationship with Peers	2.643	1.245	4.00	.275	-1.130	3.105	1.344	4.00	-.060	-1.309

Table 6.

Intercorrelations among Scales, and between Scales and Final Course Grade at T1 and T2.

	1	2	3	4	5	6	7	8	9	10	11	12
Self-System Processes												
1. Competence	--	.346	.645	.121	.323	-.071[ns]	-.351	.703	.367	.271	.125	.400
2. Autonomy	.414	--	.513	.511	.619	-.473	-.530	.463	.390	.266	.270	.252
3. Relatedness	.628	.550	--	.314	.482	-.316	-.505	.703	.490	.290	.341	.352
Engagement vs Disaffection												
4. Behavioral Engagement	.204	.475	.370	--	.420	-.673	-.298	.179	.146	.224	.255	.240
5. Emotional Engagement	.352	.617	.506	.439	--	-.341	-.672	.342	.176	.226	.301	.229
6. Behavioral Disaffection	-.181	-.477	-.379	-.647	-.378	--	.477	-.200	-.191	-.178	-.266	-.176
7. Emotional Disaffection	-.436	-.520	-.501	-.302	-.620	.480	--	-.352	-.164	-.213	-.192	-.279
Identity as a Scientist												
8. Science Identity	.695	.498	.740	.267	.401	-.247	-.403	--	.586	.351	.153	.326
9. Science Career Plans	.309	.359	.514	.178	.212	-.178	-.155	.552	--	.210	.175	.128*
10. Purpose in Science	.271	.337	.350	.302	.269	-.240	-.201	.413	.269	--	.126	.128*
Positive Relationship (11)	.130	.210	.344	.138	.271	-.100	-.143	.164	.193	.078*	--	.151*
Final Course Grade (12)	.194	.153	.140	.196	.157	-.177	-.184	.148	-.060[ns]	.110*	.069[ns]	--

Note. T1 N = 856; T2 N = 574; T1 correlations below diagonal. T2 correlations above diagonal. All correlations are significant at $p < .01$ unless otherwise noted; * $p < .05$; [ns] = non-significant.

Figure Caption

Figure 1. A Self-Determination theoretical model of motivational processes in undergraduate science classes, in which (1) students' previous experiences (including preparation, performance, and motivation in science) as well as (2) the nature of the academic work they encounter in class and (3) the supportiveness of their current relationships with instructors and peers, together predict (4) their motivation in science, including a sense of competence, autonomy, and relatedness, identity as a scientist, and engagement in science learning. In turn, these motivational dispositions, which are themselves interconnected, shape (5) students' success and persistence in STEM coursework, which then (6) feeds back into their subsequent motivation for science.

Appendix

Items from the Self-determination, Purpose, Identity, and Engagement in Science (SPIRES) survey.

Self-System Processes

1. **Competence**

- I am good at science. (+)
- I find it easy to understand the things we are learning in this class. (+)
- Even if they are challenging, I can do well in my science classes. (+)
- I don't have the intelligence/brains to succeed in science. (-)
- When I do poorly in a science course, I usually can't figure out why. (-)

2. Autonomy/Ownership

Why do I do my classwork and homework for this course?

- Because I want to understand the subject. (+)
- Because I want to learn new things. (+)
- Because doing well in science is important to me. (+)
- Because it's fun to answer challenging science questions. (+)
- For this class, I just learn the stuff I have to in order to pass the test(s). (-)

3. Relatedness/Belonging

- This course is a good place for students like me. (+)
- This is the right course for me to be taking now. (+)
- In science courses, I feel like an outsider. (-)
- I fit in well with the other students in this class. (+)
- I'm not really sure that science is the right major is for me. (-)

Engagement vs. Disaffection

4. Behavioral Engagement

- I pay attention in class.
- I study for this class.
- I try hard to understand the professor's lectures.
- I keep up with the work for this class.
- I try hard to do well in this class.

5. Emotional Engagement

- I enjoy the time I spend in this class.
- The material we cover is interesting.
- It's exciting to make connections between the ideas learned in this class.
- The material we cover in class is challenging (in a good way).
- The readings for this class are interesting

6. Behavioral Disaffection

- It's hard to make myself come to this class.
- Outside of class, I don't put much work in on this course.

- Anything I do for this class is always last minute.
- I don't really study for this class.
- I work on other things when I'm in this class.

7. Emotional Disaffection

- When in class, I feel bored.
- This class is stressing me out.
- This class can be pretty dull.
- When I'm in this class, I can't wait for it to be over.
- This class is no fun.

Identity as a Scientist

8. Science Identity

- I am the kind of person who can succeed in science. (+)
- I think that science is fascinating. (+)
- I feel at home in science. (+)
- Sometimes I feel like I don't belong in science. (-)
- I don't think I could ever really feel comfortable in science. (-)

9. Science Career Plans

- For the career I want, I need a degree in science. (+)
- I am planning on a job that involves science. (+)
- Science is important for my future career. (+)
- I'm just not cut out for a career in science. (-)

10. Purpose in Science

- Science can help solve many of society's problems. (+)
- I believe that science can help make the world a better place. (+)
- I can see lots of ways that science makes a positive difference in our everyday lives. (+)
- If everyone in our society learned more about science, we could all make better decisions about important things like politics, medicine, and the environment. (+)

11. Positive relationships and collaborations

- I have gotten to know other students in this class. (+)
- In this class, I have found people to study with. (+)
- In this class, I know people I could ask for help with assignments. (+)
- Some students from this class and I are thinking about taking another course together. (+)

Table S1. Fit statistics for single-factor models for each scale at T1 and T2.

Model	Time Point	χ^2 (*df*)	CFI	TLI	RMSEA (90% CI)	SRMR
Self-system Processes						
Competence	T1	26.645*** (5)	.977	.954	.071 (.046 - .099)	.032
	T2	55.969*** (5)	.931	.862	.133 (.103 - .166)	.052
Autonomy	T1	51.363*** (5)	.968	.936	.104 (.079 - .131)	.031
	T2	65.475*** (5)	.952	.904	.145 (.115 - .177)	.035
Relatedness	T1	42.644*** (5)	.949	.897	.094 (.069 - .121)	.040
	T2	34.651*** (5)	.950	.899	.102 (.071 - .135)	.044
Engagement vs. Disaffection						
Behavioral Engagement	T1	44.352*** (5)	.959	.918	.096 (.071 - .123)	.036
	T2	41.964*** (5)	.951	.902	.113 (.083 - .146)	.042
Emotional Engagement	T1	29.970*** (5)	.982	.964	.076 (.051 - .104)	.025
	T2	24.826*** (5)	.981	.961	.083 (.052 - .117)	.026
Behavioral Disaffection	T1	25.548*** (5)	.971	.942	.069 (.044 - .097)	.029
	T2	17.466** (5)	.984	.968	.066 (.034 -.101)	.028
Emotional Disaffection	T1	15.752** (5)	.991	.983	.050 (.023 - .079)	.020
	T2	21.555** (5)	.987	.973	.076 (.045 - .110)	.023

Identity as a Scientist						
Science Identity	T1	67.203*** (5)	.955	.911	.121 (.096 - .147)	.043
	T2	47.631*** (5)	.959	.917	.122 (.092 - .155)	.038
Science Career Plans	T1	8.631* (2)	.995	.985	.062 (.024 - .107)	.019
	T2	11.140** (2)	.992	.975	.089 (.043 - .143)	.025
Purpose in Science	T1	14.592** (2)	.995	.984	.086 (.048 - .129)	.013
	T2	.481 ns (2)	1.000	1.002	.000 (.000 - .053)	.002
Positive Relationships with Peers	T1	6.403* (2)	.997	.992	.051 (.009 - .097)	.011
	T2	4.797 ns (2)	.998	.994	.049 (.000 - .108)	.010

Note. T1 $N = 856$; T2 $N = 574$.

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

Table S2. Item loadings (range and average) and number of correlations residuals > .10 for all scales at T1 and T2.

		Loadings		Correlation Residuals
	Time Point	Range	Average	Number >.10
Self-System Processes				
Competence	T1	.324 -.783	.600	0
	T2	.320 - .859	.607	2
Autonomy	T1	.393 - .868	.671	0
	T2	.465 - .918	.711	1
Relatedness	T1	.477 - .745	.576	1
	T2	.434 - .827	.587	2
Engagement vs Disaffection				
Behavioral Engagement	T1	.538 - .738	.624	0
	T2	.532 - .751	.642	1
Emotional Engagement	T1	.566 - .826	.690	0
	T2	.665 - .802	.713	0
Behavioral Disaffection	T1	.262 - .727	.549	0
	T2	.356 - .822	.626	0
Emotional Disaffection	T1	.227 - .807	.626	0
	T2	.228 - .840	.698	0
Identity as a Scientist				
Science Identity	T1	.620 - .803	.687	1
	T2	.640 - .779	.709	1
Science Career Plans	T1	.419 - .931	.712	0
	T2	.417 - .960	.737	1
Purpose in Science	T1	.786 - .919	.850	0
	T2	.858 - .912	.891	0
Positive Relationships with Peers	T1	.648 - .905	.788	0
	T2	.772 - .906	.838	0

Table S3. Fit statistics and results of $\Delta\chi^2$ analyses for the configural, metric, and scalar invariance models for the scales tapping the three self-system processes.

Construct	Measurement Invariance Model	χ^2 (*df*)	CFI	TLI	RMSEA (90% CI)	SRMR	$\Delta\chi^2$ (Δ *df*)	ΔCFI
Self-System Processes								
Competence	Configural	43.168*** (15)	.970	.941	.081 (.053 - .110)	.034	---	----
	Metric	50.000** (23)	.972	.963	.064 (.040 - .088)	.040	6.833^{ns} (8)	.002
	Scalar	80.748*** (31)	.948	.949	.075 (.055 - .095)	.052	30.747*** (8)	-.024
Autonomy	Configural	52.052*** (15)	.975	.950	.093 (.066 - .121)	.027	---	---
	Metric	72.478*** (23)	.966	.956	.087 (.065 - .110)	.049	20.426** (8)	-.009
	Scalar	102.083*** (31)	.952	.953	.090 (.071 - .109)	.058	29.605*** (8)	-.014
Relatedness	Configural	59.028*** (15)	.940	.881	.101 (.075 - .129)	.042	---	---
	Metric	68.735*** (23)	.938	.919	.083 (.061 - .107)	.050	9.708^{ns} (8)	-.002
	Scalar	83.832*** (31)	.929	.931	.077 (.058 - .097)	.055	15.096^{ns} (8)	-.009

Note. T1 data were used for all measurement invariance models; ***$p < .001$; ** $p < .01$; * $p < .05$

Table S4. Fit statistics and results of $\Delta\chi^2$ analyses for the configural, metric, and scalar invariance models for the four engagement versus disaffection survey scales.

Construct	**Measurement Invariance Model**	χ^2 (*df*)	**CFI**	**TLI**	**RMSEA (90% CI)**	**SRMR**	$\Delta\chi^2$ (Δ *df*)	**ΔCFI**
Behavioral Engagement	Configural	57.686*** (15)	.956	.912	.100 (.073 - .128)	.034	---	---
	Metric	64.741*** (23)	.957	.944	.080 (.057 - .103)	.043	7.056^{ns} (8)	.001
	Scalar	78.401*** (31)	.951	.953	.073 (.053 - .094)	.048	13.659^{ns} (8)	-.006
Emotional Engagement	Configural	46.531*** (15)	.977	.953	.086 (.059 - .114)	.027	---	---
	Metric	58.508*** (23)	.974	.966	.074 (.050 - .097)	.043	11.976^{ns} (8)	-.003
	Scalar	92.664*** (31)	.954	.956	.083 (.064 -.103)	.057	34.157*** (8)	-.02
Behavioral Disaffection	Configural	44.407*** (15)	.960	.920	.083 (.055 - .112)	.036	---	---
	Metric	72.310*** (23)	.933	.913	.087 (.065 - .110)	.056	27.903*** (8)	-.027
	Scalar	78.517*** (31)	.935	.938	.073 (.053 - .094)	.058	6.207^{ns} (8)	.002
Emotional Disaffection	Configural	22.986^{ns} (15)	.994	.987	.043 (.000 - .076)	.020	---	---
	Metric	35.864* (23)	.990	.986	.044 (.008 - .071)	.037	12.879^{ns} (8)	-.004
	Scalar	62.452** (31)	.974	.975	.060 (.038 - .081)	.054	26.587*** (8)	-.016

Note. T1 data were used for all measurement invariance models; ***$p < .001$; ** $p < .01$; * $p < .05$

Table S5. Fit statistics and results of $\Delta\chi^2$ analyses for the configural, metric, and scalar invariance models for the three science identity survey scales and the Positive Relationships with Peers scale.

Construct	Measurement Invariance Model	χ^2 (*df*)	CFI	TLI	RMSEA (90% CI)	SRMR	$\Delta\chi^2$ (Δ *df*)	ΔCFI
Identity as a Scientist								
Science Identity	Configural	81.852*** (15)	.952	.904	.125 (.099 - .152)	.040	---	---
	Metric	89.429*** (23)	.952	.938	.101 (.079 - .123)	.049	7.577 ^{ns} (8)	.000
	Scalar	98.771*** (31)	.951	.953	.088 (.068 - .107)	.051	9.342 ^{ns} (8)	-.001
Science Career Plans	Configural	10.660^{ns} (6)	.996	.989	.052 (.000 - .102)	.015	---	---
	Metric	22.176* (12)	.992	.988	.055 (.014 - .090)	.039	11.515^{ns} (6)	-.004
	Scalar	25.318^{ns} (18)	.994	.994	.038 (.000 - .069)	.041	3.143^{ns} (6)	.002
Purpose in Science	Configural	27.185*** (6)	.991	.974	.111 (.071 - .155)	.013	---	----
	Metric	55.243*** (12)	.983	.974	.112 (.083 - .143)	.058	28.058*** (6)	-.008
	Scalar	61.203*** (18)	.983	.983	.092 (.067 - .117)	.059	5.961^{ns} (6)	.000
Positive Relationships with Peers	Configural	9.455^{ns} (6)	.998	.994	.045 (.000 - .097)	.010	---	---
	Metric	26.422** (12)	.992	.988	.065 (.031 - .099)	.041	16.967** (6)	-.006
	Scalar	54.420*** (18)	.979	.979	.084 (.059 - .110)	.054	27.998*** (6)	-.013

Note. T1 data were used for all measurement invariance models; ***p < .001; ** p < .01; * p <.05

Appendix

Items from the Self-determination, Purpose, Identity, and Engagement in Science (SPIRES) survey.

Self-System Processes

1. **Competence**
   - I am good at science. (+)
   - I find it easy to understand the things we are learning in this class. (+)
   - Even if they are challenging, I can do well in my science classes. (+)
   - I don't have the intelligence/brains to succeed in science. (-)
   - When I do poorly in a science course, I usually can't figure out why. (-)
2. **Autonomy/Ownership**
   Why do I do my classwork and homework for this course?
   - Because I want to understand the subject. (+)
   - Because I want to learn new things. (+)
   - Because doing well in science is important to me. (+)
   - Because it's fun to answer challenging science questions. (+)
   - For this class, I just learn the stuff I have to in order to pass the test(s). (-)
3. **Relatedness/Belonging**
   - This course is a good place for students like me. (+)
   - This is the right course for me to be taking now. (+)
   - In science courses, I feel like an outsider. (-)
   - I fit in well with the other students in this class. (+)
   - I'm not really sure that science is the right major is for me. (-)

Engagement vs. Disaffection

4. **Behavioral Engagement**
   - I pay attention in class.
   - I study for this class.
   - I try hard to understand the professor's lectures.
   - I keep up with the work for this class.
   - I try hard to do well in this class.
5. **Emotional Engagement**
   - I enjoy the time I spend in this class.
   - The material we cover is interesting.
   - It's exciting to make connections between the ideas learned in this class.
   - The material we cover in class is challenging (in a good way).
   - The readings for this class are interesting
6. **Behavioral Disaffection**
   - It's hard to make myself come to this class.
   - Outside of class, I don't put much work in on this course.

- Anything I do for this class is always last minute.
- I don't really study for this class.
- I work on other things when I'm in this class.

7. Emotional Disaffection

- When in class, I feel bored.
- This class is stressing me out.
- This class can be pretty dull.
- When I'm in this class, I can't wait for it to be over.
- This class is no fun.

Identity as a Scientist

8. Science Identity

- I am the kind of person who can succeed in science. (+)
- I think that science is fascinating. (+)
- I feel at home in science. (+)
- Sometimes I feel like I don't belong in science. (-)
- I don't think I could ever really feel comfortable in science. (-)

9. Science Career Plans

- For the career I want, I need a degree in science. (+)
- I am planning on a job that involves science. (+)
- Science is important for my future career. (+)
- I'm just not cut out for a career in science. (-)

10. Purpose in Science

- Science can help solve many of society's problems. (+)
- I believe that science can help make the world a better place. (+)
- I can see lots of ways that science makes a positive difference in our everyday lives. (+)
- If everyone in our society learned more about science, we could all make better decisions about important things like politics, medicine, and the environment. (+)

11. Positive relationships and collaborations

- I have gotten to know other students in this class. (+)
- In this class, I have found people to study with. (+)
- In this class, I know people I could ask for help with assignments. (+)
- Some students from this class and I are thinking about taking another course together. (+)