Title
Utility value trajectories and their relationship with behavioral engagement and performance in introductory statistics

Authors
Sutter, Claudia C
Hulleman, Chris S
Givvin, Karen B
et al.

Publication Date
2022

DOI
10.1016/j.lindif.2021.102095

Learning and Individual Differences 93 (2022) 102095

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Utility value trajectories and their relationship with behavioral engagement and performance in introductory statistics

Claudia C. Sutter [a,b,*], Chris S. Hulleman [a], Karen B. Givvin [b], Mary Tucker [b]

[a] School of Education and Human Development, University of Virginia, United States of America
[b] Department of Psychology, University of California, Los Angeles, United States of America

ARTICLE INFO

Keywords:
Utility value
Behavioral engagement
Introductory statistics
Gender
Race

ABSTRACT

This study examined utility value trajectories overall and by gender, race, and underrepresented racial minority (URM) status within an introductory statistics course and tested the relationships between utility value, behavioral engagement, and performance. Data from 1108 undergraduates included three surveys integrated into their online textbook (t_1: beginning of the textbook; t_2: middle of the textbook; t_3: end of the textbook). On average, utility value declined from t_1 to t_2. There were no significant differences by gender, however, latent change models revealed significant differences between URM and non-URM students: While Black, Latinx, and racially minoritized students continued to experience a decline in utility value from t_2 to t_3, White and Asian students did not. Utility value was reciprocally related to behavioral engagement during the learning process, and both utility value (t_3) and behavioral engagement (t_1 and t_2) predicted final course grades. The findings highlight the need for a deeper understanding of short-term relationships between utility value, behavioral engagement, and performance as well as the ongoing concern for how best to support students who identify with underrepresented groups in STEM.

1. Introduction

A major challenge and concern in the US is bridging the theory-practice gap by making learning statistics more useful and relevant to students' lives by empowering them with statistical skills that can be applied beyond the classroom (Songsore & White, 2018). Introductory statistics courses not only serve as gateway courses for students majoring in STEM, but also for students pursuing other majors and careers. For instance, statistical training is often considered a requirement for Psychology undergraduate majors and an important component of the curriculum (Friedrich et al., 2000). Statistics not only plays "a central role in helping students achieve many of the critical thinking goals outlined in the American Psychological Association", but its advances have been "identified as part of psychology's many significant contributions to work carried out in a range of STEM fields" (Friedrich et al., 2018, p. 312).

Whether students have positive experiences, are motivated, and perceive what they are learning as useful and valuable plays a critical role during college and can not only determine whether they persist in their chosen major, but also whether students apply statistics in their everyday lives (Kosovich, Flake, et al., 2017; Rosenzweig et al., 2019). A single course in college can be the deciding point not only between getting a degree or not, but between continuing to pursue a career in their chosen field (Goudas & Boylan, 2012). However, we know little about how students' motivation changes over the duration of a single course in introductory statistics and how it relates to short-term learning processes and outcomes, because most prior research spans over long periods of time (Kosovich, Flake, et al., 2017). In particular, we know little about how motivational trajectories may differ among different subgroups of students, including by gender or race. Prior research in STEM points to potential differential motivational trajectories between female and male students as well as racially minority and majority students. For instance, female and racially minoritized students are particularly at risk for facing structural barriers in college, especially in STEM, and colleges may unintentionally send messages of non-belonging to underrepresented students, which in turn can carry important implications for academic outcomes (e.g., Strayhorn, 2012). Thus, gaining a deeper understanding of how motivation relates to engagement and learning specifically in the context of statistics courses will enable us to better design learning contexts and opportunity structures to support students from traditionally marginalized and minoritized backgrounds (e.g., Black and Latinx students; Gray et al., 2018).

* Corresponding author at: University of California, Los Angeles, Department of Psychology, 1285 Psychology Building, Box 951563, Los Angeles, CA 90095.
E-mail address: csutter@psych.ucla.edu (C.C. Sutter).

https://doi.org/10.1016/j.lindif.2021.102095
Received 18 March 2021; Received in revised form 21 October 2021; Accepted 1 November 2021
Available online 3 December 2021

As part of a math requirement, many students take a statistics course at some point during their undergraduate studies. However, few undergraduates are required to take more than one statistics course. As such, there is a very narrow window during which to examine students' motivation, even as introductory statistics courses are likely to shape students' attitudes toward the topic in the longer term. Students who leave statistics courses with negative attitudes are unlikely to apply what they have learned, whereas students who believe statistics is useful will be more likely to engage in statistical tasks, apply what they have learned, and enroll in future statistics courses (Schau & Emmioğlu, 2012). This is especially important in majors where students may only take one statistics course, like Psychology.

The purpose of this study is to examine students' motivation as they complete an initial course on statistics and data science. Specifically, the aim is to investigate the trajectories of students' perceived utility value (i.e., perceptions of the usefulness and relevance of course material to their lives) and its association with engagement and performance as students progress through the course. We focus on utility value for two reasons. From a theoretical perspective, we want to understand how students' perceived utility value of the course develops and relates to students' engagement in the course material and their achievement. From a practical perspective, we want to identify intervention opportunities in terms of how we could address the well documented decline in student's motivation over time (Jacobs et al., 2002) by focusing on utility value, "which seems the most amenable to a classroom intervention (...) given its more external nature" (Hulleman et al., 2010, p. 891). Given the importance of statistics for both STEM and non-STEM majors as well as the particular concern with broadening STEM participation among female and racially minoritized students, we further examined motivational change considering students' gender and underrepresented racial minority status (i.e., students from African American, Latinx, or other racially minoritized groups).

2. Theoretical background

2.1. Perceived utility value

Perceiving what they are learning as useful and valuable is a crucial predictor of academic engagement and achievement for all students, especially in the higher education environment (Kosovich, Flake, et al., 2017). Utility value has been found to be a particularly important predictor of student learning outcomes in STEM such as motivation and performance (Gaspard et al., 2015; Hulleman et al., 2008, 2010). According to the utility-value process model, when students perceive the material to be useful to their lives and future goals, they are subsequently more likely to engage in it, feel more autonomously motivated, and identify with the course content (Hulleman, 2007). However, significantly less is known about how utility value changes over time, how changes are related to engagement and learning outcomes, and how student backgrounds and contextual factors are related to utility value (Bathgate & Schunn, 2017).

One of the most comprehensive theoretical frameworks for understanding the role of values in education is the expectancy-value framework of achievement motivation (Eccles et al., 1983), which has been updated to the expectancy-value-cost framework to distinguish the unique role of cost (Barron & Hulleman, 2015). This framework proposes that achievement and achievement-related choices and behaviors are most proximally determined by expectations of success, perceptions of value for the task, and perceived costs to engaging in the task. In particular, numerous studies have provided evidence for the association between values and choice related behaviors in STEM (e.g., Guo et al., 2015; Perez et al., 2014; Wigfield & Eccles, 2002), as well as interest, persistence, and performance outcomes (e.g., Durik et al., 2006; Fong et al., 2021; Hulleman et al., 2010). In this study we focus on the role that students' perceptions of the utility value of the material play in their engagement and learning in the course.

Understanding motivation for statistics in general, and the role of perceived utility value in particular, is increasingly important in higher education as statistics courses take on an increasing role in satisfying quantitative reasoning requirements in higher education (Bateiha et al., 2020; Chiesi & Primi, 2010; Ramirez et al., 2010). Given that many students enter such courses simply to satisfy a requirement, which most often results in lower levels of motivation than their major courses (Kosovich, Hulleman, et al., 2017), understanding what leads students to see meaning and personal relevance of these courses to their lives is an important initial indicator of persistence and learning. Although there is a relative dearth of research using the expectancy-value-cost framework to understand levels of student motivation in statistics courses, some initial research suggests that perceived value for statistics is positively related to effort, persistence, and performance in introductory statistics in college (e.g., Hood et al., 2012; Schau et al., 1995; Sorge & Schau, 2002).

2.1.1. Motivation for all or motivation for some?

STEM learning environments often result in lower motivation, persistence, and achievement for students who identify with traditionally marginalized and minoritized groups, including women, Black, Latinx, and Indigenous American students, and students from lower socioeconomic status (Eccles, 2007; Fong et al., 2021; Harris et al., 2020; Riegle-Crumb et al., 2019; Wang & Degol, 2013). The contextual factors that contribute to these differences include societal stereotypes about race and gender, curriculum that emphasizes more masculine and Caucasian values, and the inability to support students who enter the classroom with differing levels of academic preparation (Rosenzweig & Wigfield, 2016). In order to better understand how we can support minoritized students in successfully completing STEM courses, we need to explore the motivational experiences among female and underrepresented racially minoritized students. The findings regarding gender differences in how much students value STEM domains or courses are mixed, with some studies pointing to significant gender differences in utility value favoring boys (Gaspard et al., 2015) and others revealing no significant differences by gender in levels of STEM value (Eccles, 2009). Studies focused on math continue to show gender differences in self-perceptions, self-concept, and success expectancies in STEM fields favoring male students (Gaspard et al., 2015; Jacobs et al., 2002). The few studies conducted in introductory statistics show that, despite similar levels of value toward statistics (for an overview see Ramirez et al., 2010), female students feel less confident in their ability and have fewer positive attitudes toward statistics than their male counterparts (Van Es & Weaver, 2018).

Although it is well documented that African American and Latinx/Hispanic students are historically underrepresented in STEM fields in general, less research has examined racial differences in attitudes toward STEM and even less research on statistics (Wiebe et al., 2018). The few studies that exist report inconsistent findings, and the variety of racial groups, age groups, and measurement instruments makes it difficult to make generalizations. Therefore, our study will contribute to prior research by examining differences in utility value, behavioral engagement and learning outcomes in statistics according to students' gender and racial background.

2.2. Change in motivation and utility value over time

Value beliefs and similar motivational constructs in STEM are known to decline in the long-term, especially as students progress through primary and secondary school (Jacobs et al., 2002; Watt, 2004). Over the past decade, research on motivational change at the postsecondary level has gained momentum with studies reporting declining levels motivation in introductory STEM courses such as in chemistry (Young et al., 2018; Zusho et al., 2003), biology (Gibbens, 2019; Rybczynski & Schussler, 2013; Young et al., 2018), engineering (Robinson et al., 2019), as well as physics and mathematics (Musu-Gillette et al., 2015;

Young et al., 2018), with particular declines in perceived usefulness (Rybczynski & Schussler, 2013). Although studies have shown that students' attitudes and their motivation in statistics courses change in the short-term (i.e., from the beginning to the end of introductory courses; Rhoads & Hubele, 2000; Schau & Emmioğlu, 2012), the trajectories of utility value within the context of a single introductory statistics course are less clear. Because motivation has been shown to vary between content and subject areas (Crede & Phillips, 2011), additional research is needed to gain an understanding of the short-term change in perceived usefulness in statistics.

While research based on expectancy-value-cost motivation has proven effective in explaining differences in achievement-related outcomes such as academic and career choices based on gender, race, and ethnicity (Eccles, 2005, 2009; Wigfield et al., 2009), there are relatively few studies that specifically examine motivational *change* across demographic subgroups. The limited research on motivational change in STEM courses across gender and race report somewhat inconsistent findings: Some studies observed no significant differences in motivational change by gender or by underrepresented or minoritized status (Kosovich, Flake, et al., 2017; Robinson et al., 2019; Young et al., 2018), whereas other studies report differences by gender (Chouinard & Roy, 2008; Fredricks & Eccles, 2002). Overall, findings across demographic subgroups are inconclusive and have primarily focused on the motivational development among primary and secondary school students. Hence, examining motivational trajectories differentiated by gender and racial minority status at the college level is in need of further examination as it provides a unique environment with regard to the motivational pattern and its association with learning processes and outcomes (Robinson et al., 2019). Assessing short-term change in motivation within a single introductory statistics course is crucial, because instructors can impact how student motivation changes throughout the semester (Bathgate & Schunn, 2017; Young et al., 2018). A deeper understanding of individual differences in motivational change will help identify targets and inform the timing and design of future motivational interventions in STEM courses.

2.3. Utility value, behavioral engagement, and performance

From an expectancy-value-cost perspective, students' motivation in general - and perceived utility value in particular - is closely linked to persistence and effort with which they pursue those tasks and activities, and their task performance (Wigfield et al., 2015). In addition to the numerous correlational studies that show the relationship between perceived utility value and outcomes such as course grades, test performance, persistence, and subject-matter interest (e.g., Durik et al., 2006; Eccles & Wigfield, 2002; Hulleman et al., 2010, 2017), there is a burgeoning research literature showing that utility-value interventions are effective at boosting learning, performance, and motivation in a variety of STEM disciplines (Rosenzweig & Wigfield, 2016), including statistics (Acee & Weinstein, 2010; Kosovich et al., 2019). Specifically, utility value has been shown to increase pass rates (e.g., in intermediate algebra courses; Kosovich et al., 2019), and course grades (e.g., in college psychology; Hulleman et al., 2010). Motivation more generally is also linked to students' choices about which learning tasks and activities to engage in. Although there is no single definition of academic engagement, there is consensus that engagement is a multidimensional construct that includes affective, behavioral, and cognitive components (Fredricks et al., 2014; Martin, 2007; Skinner et al., 2008, 2009). At the core of numerous motivational conceptualizations, engagement from a behavioral lens captures the quality of an individual's involvement and participation in learning tasks and activities as well as their connection and interaction with the learning materials within the learning environment (Skinner et al., 2008).

The expectancy-value-cost framework (Barron & Hulleman, 2015) suggests that perceptions of task value directly impact choice, persistence, and performance, with engagement acting as a translator of motivation into action (Wigfield et al., 2015). For example, students who recognize the importance and usefulness of the content of a course or an activity will most likely devote more effort, persist more, and perform better when they engage in a task of activity. Achievement-related choices, in turn, predict task values (e.g., utility value): Students' engagement from a behavioral perspective (i.e., engaging with the learning material and participating in learning activities) provides the learner with opportunities to connect more deeply with the content, thus, supporting them in finding the value and relevance in an activity (Bathgate & Schunn, 2017; Durik et al., 2006). Thus, utility value and behavioral engagement are suggested to be in a bi-directional relationship.

Although numerous instruments have been developed to assess engagement, two concerns related to the operationalization persist: (a) There is considerable variability in how engagement in the academic context is operationalized, with the majority of research relying on self-report measures, experience sampling methods, or observations and (b) the validity of these instruments is often unknown (Fredricks et al., 2014).

In this study, we did not seek to resolve this tension. Instead, we adopted a practical measurement approach (Kosovich, Hulleman, et al., 2017; Krumm et al., 2016) focusing on behavioral engagement by using authentic classroom tasks. The advantages of existing classroom tasks are the reduced response burden on students, the value to students in completing a task that is part of the course, and the honoring of assessments that have value to practitioners. The downside is that these measures were not developed by measurement experts, and therefore may not meet traditional standards of validity and reliability. Based on the conceptualization of behavioral engagement, which "draws on the idea of participation and includes involvement in academic, social, or extracurricular activities and is considered crucial for achieving positive academic outcomes and preventing dropping out" (Fredricks & McColskey, 2012, p. 764), we used students' participation, more specifically the completion of, and performance on, voluntary end-of-chapter review questions as our measure of behavioral engagement.

3. The present study

The present study examined changes in perceived utility value within a single introductory statistics course and tested the relationship between utility value, behavioral engagement, and achievement. Specifically, we sought to address the following research questions:

1. What are students' initial utility value levels in an introductory statistics class and how do they change over the course of the class?
2. Are there differences in the levels of and changes in utility value among subgroups of students (e.g., differentiated by gender; race; racially minoritized status)?
3. What is the relationship between utility value, behavioral engagement in the learning process, and course grades (controlling for prior GPA, gender, and racially minoritized status)?

First, based on prior research (Fredricks & Eccles, 2002; Jacobs et al., 2002; Kosovich, Flake, et al., 2017), we expected utility value to decline over the course of the introductory statistics class. Second, due to the limited number of prior research on motivational trajectories across demographic subgroups, no specific hypotheses about differential trajectories were made. Although prior research has documented lower levels of motivation for students from traditionally minoritized groups in STEM (Harris et al., 2020; Riegle-Crumb et al., 2019), recent research at the postsecondary level found no gender or race differences (Kosovich, Flake, et al., 2017; Robinson et al., 2019; Young et al., 2018) in rates of motivational change over time. Third, with prior correlational research revealing that perceptions of utility value predict engagement, effort, and academic achievement (e.g., Hulleman et al., 2008; Wigfield & Cambria, 2010), we hypothesized that utility value would be related

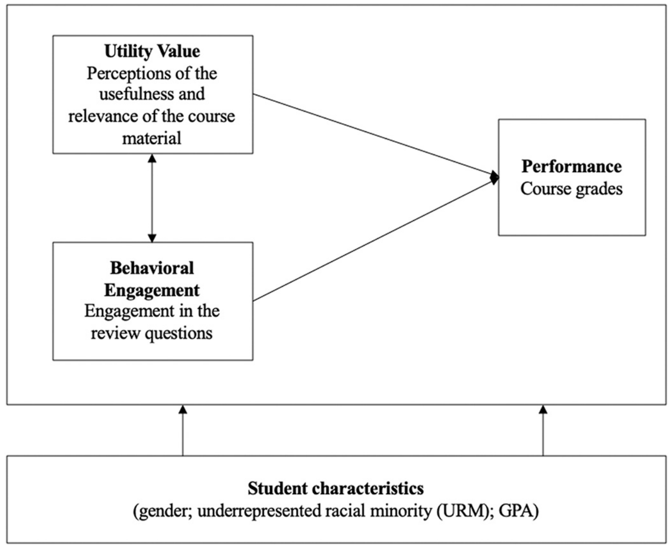

Fig. 1. Theoretical Logic Model: How students' perceived utility value and engagement in the course interrelate and predict course performance.

to behavioral engagement and outcomes (see Fig. 1).

4. Method

4.1. Participants, course setting, and procedure

Data were collected from 1018 students in six different course sections of introductory statistics at one university in California. The sample was predominantly female (71.8%), 39.2% Asian, 26.0% White, 17.0% Latinx, and 3.2% African American (3.1% did not disclose their race and 11.4% students preferred to self-describe). Of those who indicated their major, almost 90% were Psychology majors (37.6% Pre-Psychology, 34.3% Pre-Psychobiology, 14.2% Precognitive Science, and 2% Linguistics and Psychology).

All sections took place between June 2019 and September 2020 and used the same interactive online textbook, which consists of 12 chapters and 144 pages and includes over 1200 embedded formative assessments, R programming exercises, and end of chapter review activities (for a detailed overview of the course see Supplement). Data for this study were collected as part of an ongoing project, which was approved by the Institutional Review Board at the University of California, Los Angeles (IRB No: 20-001033).

Because students were nested within six course sections, we calculated intraclass correlations (ICCs) for utility value and engagement at the beginning of the course ($ICC_{Utility\_value} = 0.016$; $ICC_{Behavioral\_engagement} = 0.019$), in the middle of the course ($ICC_{Utility\_value} = 0.029$; $ICC_{Behavioral\_engagement} = 0.035$), and at the end of the course ($ICC_{Utility\_value} = 0.033$; $ICC_{Behavioral\_engagement} = 0.079$). Because these ICC values were below the level of triviality defined by Lee (2000; ICC < 0.10), and our research questions were focused on student-level indicators, multi-level models were not required.

4.2. Measures

4.2.1. Utility value

Surveys were collected at the beginning of the course (t_1; prior to chapter 1), in the middle (t_2; after chapter 8), and at end of the course (t_3; after chapter 12) and were integrated directly into the online textbook. Students' perceived utility value of the course was assessed using the average of two items ('What I learn in this course will be useful in the future' and 'The content of this course is important for me'; Kosovich et al., 2015) at the three time points ($\alpha t_1 = 0.84$; $\alpha t_2 = 0.85$; $\alpha t_3 = 0.85$). Both items were rated on a five-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree.

4.2.2. Behavioral engagement

At the end of each chapter, students were given the opportunity to complete two review activities with conceptual and application questions related to the content in each chapter. The review activities comprised multiple choice questions, open-ended response items and interactive R coding exercises that provided students with practice analyzing a new dataset. The first review activity was required as part of the homework and the second review activity was optional. Students' voluntary completion of the second review activity was used as a practical measure of behavioral engagement. A second review activity was available for chapters 2–10. To align the behavioral engagement measures with the timing of the student survey measures, students' activity 1 scores on chapter 1 reflected behavioral engagement at t_1, and then activity 2 scores were used for the remaining chapters, with scores for chapters 2–8 averaged for t_2, and scores for chapters 9 and 10 averaged for t_3.

4.2.3. Performance

Students' final letter grades in the course were obtained from the instructors and were coded using a standard GPA scale (A+ = 4.3, A = 4.0, A- =3.7, etc.).

4.2.4. Demographic variables

Students indicated their gender (male, female, non-binary) and race (African American, Asian, Latinx, White) with the option to self-describe. Due to a several students indicating specific racial/ethnic minority groups in the self-describe option, a variable inclusive of underrepresented racial minority (URM) was created to use in the analyses (0 = White and Asian[1]; 1 = African American, Black, Hispanic, Indian Subcontinent, Native American, Mixed race, or Greater Middle Eastern, e.g., Armenia, Afghanistan, Pakistan). Students of mixed race were included in the URM group, unless their race was a mix of White and

[1] Within the context of our study, we consider Asian students as a non-underrepresented racial group for the following two reasons: First, they are the largest group at the project's institution. Second, Asian and Asian American students are often positively stereotyped when it comes to academic performance (e.g., Armenta, 2010), and their academic performance often surpasses White students. Therefore, we included them as a part of the non-URM group

Asian. Students' self-reported GPA prior to starting the course (What is your GPA at this school?) was included as a covariate in all models.

4.3. Missing data

Missing data were low for utility value (between 2.8% at t1 to 7.2% at t3) and final grade (3.9%). There was no missing data for the review questions given that a lack of behavioral engagement was coded as 0. Missing values were estimated using the Full Information Likelihood Estimation (FIML) in Mplus (Muthén & Muthén, 1998–2012) for the latent change and SEM models.

4.4. Analysis

4.4.1. Descriptive analyses

In an initial step, independent sample *t*-tests were computed to test for differences between subgroups (gender, URM) in utility value, behavioral engagement, performance, and prior GPA (see Supplementary Table S1). Further, paired sample *t*-tests in SPSS (Version 27) were used to explore within-group changes between pre-, mid-, and post surveys. Effect sizes of differences were calculated using Cohen's *d*.

4.4.2. Measurement invariance

Measurement invariance has to be established to ensure that the same latent construct is being measured in the same way over time (Little, 2013; Widaman & Reise, 1997). This allows (a) for mean differences to be attributed to true change rather than measurement differences and (b) to make inferences about change over time (Newsom, 2015). Four invariance models were computed (e.g., Meredith, 1993; Newsom, 2015; Widaman & Reise, 1997). Model 1 (configural invariance) included the same factor structure over time without constraints on factor loadings or intercepts. Model 2 (metric or weak invariance) constrained the factor loadings to be equal across time points. Model 3 (strong or scalar invariance) required the factor loadings and the item intercepts to be invariant over time. Although strong invariance is sufficient for the comparison of latent means over time, we also estimated a model (Model 4, strict invariance) by additionally constraining the item residual variances to be equal over time. Measurement invariance analyses suggest strong invariance across the three time points (see Table S2 in the Supplement), with a reduction in CFI that was greater than 0.01 between the strong and strict models (Cheung & Rensvold, 2002; Little, 2013), allowing the comparison of latent means over time.

4.4.3. Latent change models - change in utility value and interactions with gender and URM

To examine change in utility value across the three time intervals and test the interaction between time and gender/URM, latent change models were computed, which measure change through latent difference variables and represent change scores corrected for random measurement error (Geiser, 2012). A neighbor change model was specified (see Fig. 2 for conceptual model; Geiser, 2012) to examine the change from the beginning to the middle of the course (Change 1) as well as the middle to the end of the course (Change 2). These changes were examined overall, and also by gender and URM, while controlling for GPA. Including gender and URM as predictors of latent changes in value allowed us to test the interaction between time and gender/URM, while accounting for measurement errors. Model fit was evaluated using the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA), and the standardized Root Mean Squared Residual (SRMR). A good level of fit is indicated when RMSEA and SRMR are less than 0.06 (acceptable fit: $\leq$ 0.08) and when CFI and TLI values exceed 0.95 (acceptable fit: between 0.90 and 0.95) (Hu & Bentler, 1999).

4.4.4. Structural equation models

To explore the relationship between utility value, behavioral engagement, and grades, a structural equation model specifying the reciprocal relationship between students' utility value to behavioral engagement, and to final course grade was estimated (see Fig. 3), controlling for gender, URM, and GPA and taking strong measurement into account. Specifically, we ran two successive models: In a first exploratory model, we specified a simple model regressing final grade on gender and URM and then in a second model we added in utility value and behavioral engagement. This allowed us to explore the variance of

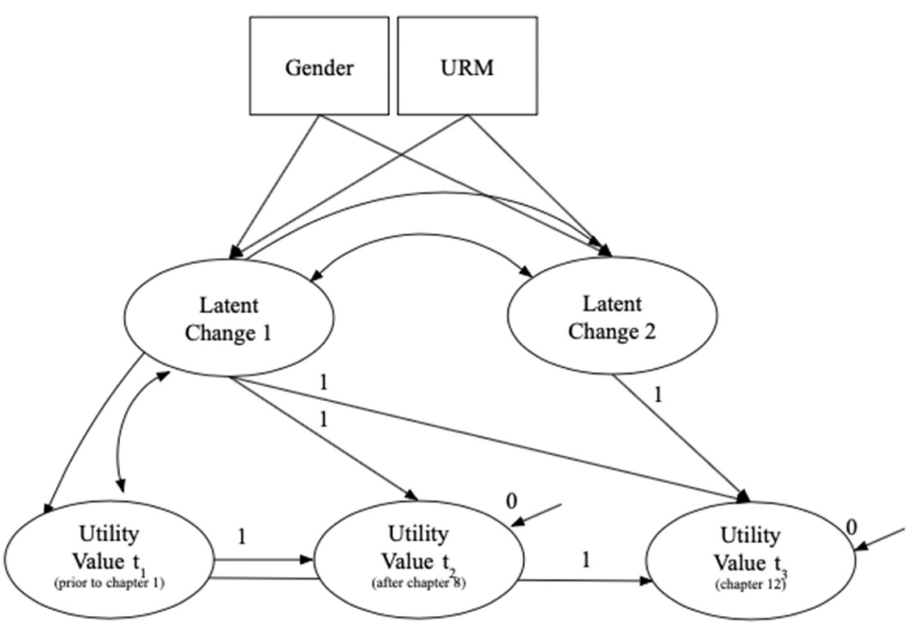

Fig. 2. Neighbor Change Model with the Proposed Effects on Students' Utility Value
Note. Change 1: from the beginning to the middle of the term; Change 2: from the middle to the end of the term. Value indicators are measured at a latent level. URM (underrepresented racial minority) includes African American, Black, Hispanic, Indian Subcontinent, Native American, Mixed race, or Greater Middle Eastern, e.g., Armenia, Afghanistan, Pakistan.

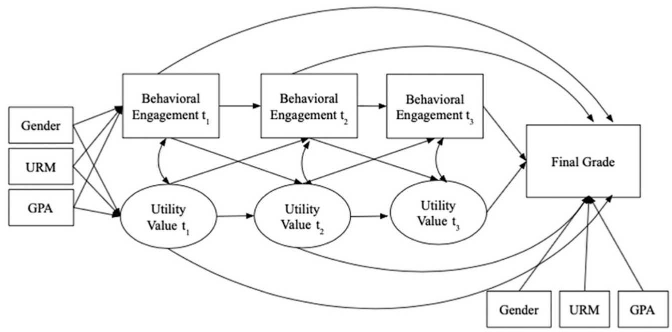

Fig. 3. Model Relating Utility Value, Behavioral Engagement, and Course Grades.
Note. Value indicators are measured at a latent level. URM (underrepresented racial minority, no = 0; yes = 1) includes African American, Black, Hispanic, Indian Subcontinent, Native American, Mixed race, or Greater Middle Eastern, e.g., Armenia, Afghanistan, Pakistan.

gender and URM gaps explained by utility value and behavioral engagement. Indirect effects were calculated via the model indirect statement in MPlus. Additionally, the total effect (sum of both indirect and direct effects) is reported (Bollen, 1987) (Fig. 3).

5. Results

5.1. Descriptive statistics

Means, standard deviations, and correlations between utility value and behavioral engagement for all three measurement occasions are reported in Table 1. Utility value at t_1 was related only to subsequent utility value but not to subsequent behavioral engagement during the course or the final course grade. However, utility value at t_2 and t_3 were significantly related to students' behavioral engagement throughout the learning process as well as the final grade. Finally, behavioral engagement was significantly related to final course grades.

Independent sample *t*-tests were computed to test for differences in the variables of interest by gender and URM (see Table S1 in Supplement). The findings revealed no significant differences in utility value or behavioral engagement by gender for the different time points. Thus, female and male students seem to equally value the course and engage in the optional course activities. However, significant differences emerged in the final course grade with male students having significantly higher grades than female students ($d = 0.16$, $p = .033$). In contrast, female students self-reported higher GPA entering the course ($d = 0.17$, $p = .032$).

Students belonging to an URM group scored significantly lower than students who identified as White or Asian in utility value at the end of the course ($d_{t3} = 0.20$), behavioral engagement throughout the term (*d's from* 0.15 to 0.49), and final grades ($d = 0.67$). URM students also reported lower GPA grades entering the course ($d = 0.41$, $p \leq .001$).

Results of paired sample *t*-tests and Cohen's *d* calculations (see Table 2) revealed that all demographic subgroups experienced a

Table 1
Descriptive statistics and correlations.

	n	Mean	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Values t_1	990	4.23	0.76	1						
(2) Values t_2	968	3.92	0.96	0.44**	1					
(3) Values t_3	945	3.89	0.95	0.34**	0.56**	1				
(4) Engagement t_1	1018	0.85	0.14	0.00	0.05	0.11**	1			
(5) Engagement t_2	1018	0.73	0.22	0.04	0.19**	0.24**	0.25**	1		
(6) Engagement t_3	1018	0.39	0.34	0.03	0.21**	0.21**	0.13**	0.34**	1	
(7) Final course grade	968	3.52	0.78	−0.01	0.12**	0.24**	0.21**	0.52**	0.27**	1

** Correlation is significant at the 0.01 level (2-tailed).

Table 2
Statistical values (t, df, p) of t-test for paired samples (t_1-t_2 and t_2-t_3) of utility value and Cohen's d (d) for the overall sample as well as by gender, race, and URM.

	Change t_2–t_1						Change t_3–t_2					
	n	t_1 *M (SD)*	t_2 *M (SD)*	*t(df)*	*p*	*d*	*n*	t_2 *M (SD)*	t_3 *M (SD)*	*t(df)*	*p*	*d*
Overall	944	4.23 (0.75)	3.93 (0.96)	−10.075 (943)	≤0.001	−0.35	922	3.93 (0.95)	3.89 (0.94)	−1.52 (921)	0.128	−0.04
By gender												
Female	693	4.24 (0.75)	3.94 (0.95)	−8.645 (692)	≤0.001	−0.35	674	3.94 (0.94)	3.88 (0.93)	−1.828 (673)	0.068	−0.06
Male	224	4.19 (0.75)	3.91 (0.98)	−34.425 (223)	≤0.001	−0.32	216	3.92 (0.96)	3.94 (0.98)	0.262 (215)	0.793	0.02
By race												
African American	30	4.40 (0.65)	4.08 (0.94)	−2.318 (29)	0.028	−0.40	31	4.05 (0.94)	3.85 (0.93)	−1.030 (30)	0.311	0–0.21
Asian	373	4.18 (0.74)	3.99 (0.90)	−3.775 (372)	≤0.001	−0.23	364	3.99 (0.90)	4.02 (0.87)	0.742 (363)	0.459	0.03
Latinx	159	4.27 (0.75)	3.92 (1.00)	−5.112 (158)	≤0.001	−0.40	151	3.95 (0.98)	3.76 (1.04)	−2.498 (150)	0.014	−0.19
White	256	4.25 (0.77)	3.87 (0.99)	−6.472 (255)	≤0.001	−0.43	248	3.87 (0.98)	3.82 (0.98)	−0.851 (247)	0.395	−0.05
Underrepresented racial minority[a]												
Yes	282	4.30 (0.74)	3.91 (1.00)	−7.499 (281)	≤0.001	−0.44	270	3.93 (0.98)	3.78 (0.98)	−2.516 (269)	0.012	−0.15
No	640	4.20 (0.75)	3.95 (0.95)	−6.918 (639)	≤0.001	−0.29	625	3.94 (0.94)	3.95 (0.92)	0.094 (624)	0.925	0.01

[a] Includes African American, Black, Hispanic, Indian Subcontinent, Native American, Mixed race, or Greater Middle Eastern, e.g., Armenia, Afghanistan, Pakistan.

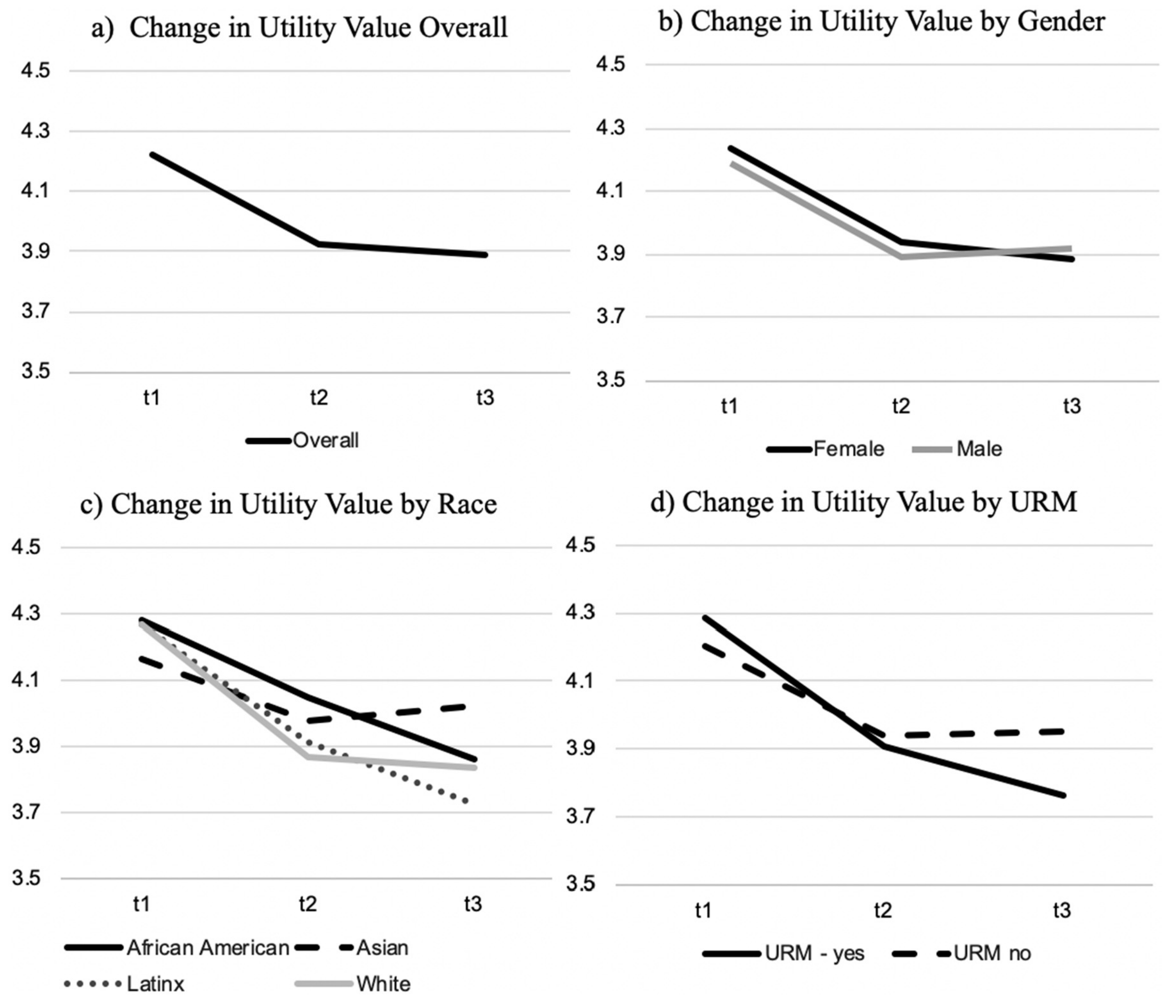

Fig. 4. a-d. Changes in Utility Value by a) Overall, b) Gender, c) Race, and d) URM (underrepresented racial minority; no = 0; yes = 1) includes African American, Black, Hispanic, Indian Subcontinent, Native American, Mixed race, or Greater Middle Eastern, e.g., Armenia, Afghanistan, Pakistan.

significant decrease in mean levels of utility value from beginning to mid-course with differential rates of decline with some groups' values decreasing more than others. Strikingly, differential patterns emerged in the development of utility value from the middle to the end of the course for different subgroups of students: Latinx ($d = -0.19$) continued to experience a significant decline in utility value toward the end of the course, whereas Asian or racial majority students did not (see Fig. 4).

5.2. *Latent change models - change in utility value and interactions with gender and URM*

A latent change model was specified to compare differences in utility value change from t_1 to t_2 and from t_2 to t_3, among subgroups of students (i.e., by gender and URM status). The model fit was good: (χ^2 (22) = 44.657, $p = .0029$; CFI = 0.992, TLI = 0.988, RMSEA = 0.033, SRMR = 0.028). The results (standardized coefficients) of the model (see Table 3) revealed that individual differences in latent change were not significantly explained by gender, indicating that the change in utility value for male and female students is not significantly different. The story was different by URM membership. Whereas the regression coefficients for the regression of Change 1 on URM was not statistically significant ($\beta = 0.059$, $p = .093$), it was for the regression of Change 2 ($\beta = 0.096$, $p = .016$) implying that there is a significant interaction between time and URM. These results align with the visual depiction of the descriptive statistics presented in Fig. 4d showing that the decline in utility value from t_1 to t_2 is similar between URM and non-URM students. However, from t_2 to t_3 URM students continue to experience a decline in utility value whereas non-URM students do not.

Table 3
Results of latent change models.

Model	Gender[a]			URM[b]		
	β	S.E.	p	β	S.E.	p
Change 1 (t_2-t_1)	0.010	0.035	0.770	−0.059	0.035	0.093
Change 2 (t_3-t_2)	−0.040	0.040	0.313	−0.096	0.040	0.016

Change 1 and Change 2 are correlated at −0.384 ($p \leq .001$).
[a] Gender (male = 0; female =1).
[b] URM (underrepresented racial minority no = 0; yes = 1) includes African American, Black, Hispanic, Indian Subcontinent, Native American, Mixed race, or Greater Middle Eastern, e.g., Armenia, Afghanistan, Pakistan.

5.3. *Structural equation model: relationship between utility value, behavioral engagement, and course grades*

Prior to exploring the relations between utility value, behavioral engagement, and achievement, we tested a model regressing gender and URM on students' final grade. The model was just-identified (χ^2 (0) = 0.000, $p \leq .001$; CFI = 1.000, TLI = 1.000, RMSEA = 0.000, SRMR = 0.000). Both gender ($\beta = -0.066$, $p = .033$) and URM status ($\beta = -0.305$, $p \leq .001$) significantly predicted final course grade. The fit of the model exploring the relationships between utility value, behavioral

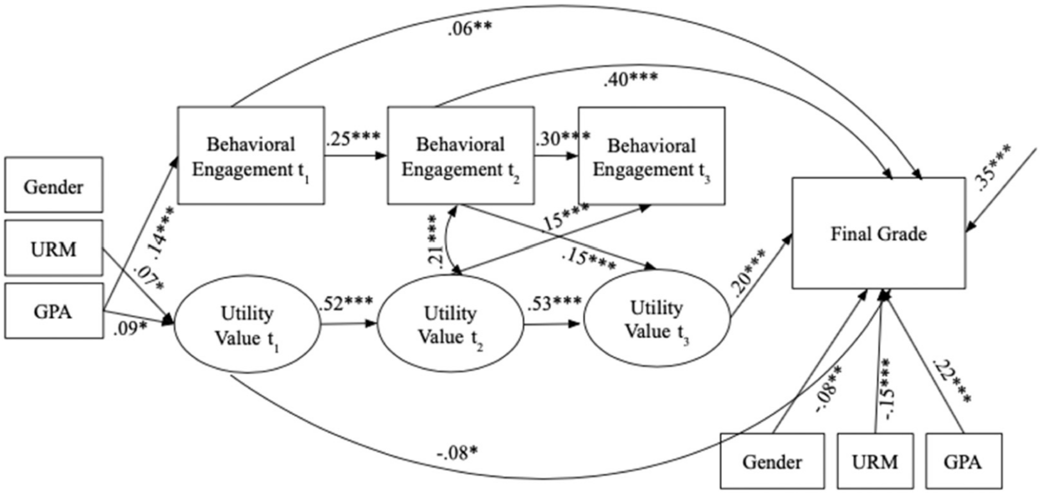

Fig. 5. Path Model of the Relationships between Utility Value, Behavioral Engagement, and Course Grades *Note. N* = 1018. Depicted values are standardized coefficients that were statistically significant (***significant at the 0.001 level; **significant at the 0.01 level; *significance at the 0.05 level). *URM* (underrepresented racial minority; no = 0; yes = 1) includes African American, Black, Hispanic, Indian Subcontinent, Native American, Mixed race, or Greater Middle Eastern, e.g., Armenia, Afghanistan, Pakistan.

engagement, and course grades (see Fig. 5) was good (χ^2 (53) = 209.336, $p \leq .001$; CFI = 0.958, TLI = 0.931, RMSEA = 0.054, SRMR = 0.052). Overall, utility value predicted subsequent utility value and behavioral engagement predicted subsequent behavioral engagement. Behavioral engagement during the course was predictive of utility value at t_3 and vice versa: Utility value at t_2 was predictive of behavioral engagement at t_3. Utility value at t_3 positively predicted final grades. Behavioral engagement at t_1 and t_2 directly predicted students' final grade. Thus, utility value at the end of the course and behavioral engagement mid-course seem particularly crucial for students' final grade. Utility value and behavioral engagement mid-course significantly related to each other. There was a small negative path relating students' initial incoming utility value with final grades, which can be explained by suppression effect (i.e., utility value at t_1 is highly correlated with utility value at t_2 and t_3 that after their shared variance is partialled out, there is nothing left to correlate with the final course grade; Cohen et al., 2002). Gender and URM status directly predicted students' final grade. For gender: the coefficient drops from −0.066 GPA points to −0.057 GPA points, which is a decrease of 0.009 GPA points. Thus, the inclusion of utility value and behavioral engagement explains 14% of the gender achievement gap. For underrepresented racial minority status, the coefficient drops from −0.305 GPA points to −0.147 GPA points, which is a decrease of 0.158 GPA points. Thus, the inclusion of utility value and behavioral engagement explains 52% of the URM achievement gap (Fig. 5).

We tested for indirect effects between utility value and behavioral engagement at t_1 and t_2 on final grade (See Table S3 in Supplement). The total indirect effect of initial utility value on final grade was not significant, however, one specific indirect effect was found via utility value at t_2 and t_3 ($\beta = 0.056, p \leq .001$). The total indirect effect of utility value at t_2 on final grades was significant, with a significant indirect effect via utility value t_3 ($\beta = 0.110, p \leq .001$), suggesting that students' perceived utility value during the course feeds into their perceived utility value toward the end of the course, which in turn relates to final course grades. The total indirect effect of behavioral engagement at t_1 was significant, with specific indirect paths via behavioral engagement at t_2; behavioral engagement at t_2 and t_3; and via behavioral engagement at t_2 and utility value at t_3. The total indirect effect of behavioral engagement at t_2 was significant with significant indirect effects via behavioral engagement and utility value at t_3.

6. Discussion

The purpose of this study was threefold: (1) to examine initial utility value levels in an introductory statistics course and how they changed across three time intervals; (2) to examine whether there are differences in how utility value changes over time by gender, race, and racially minoritized status; and (3) to explore the relationship between utility value, behavioral engagement in the learning process, and course grades.

6.1. Overall initial utility value levels and changes in utility value

Overall, initial levels of utility value were relatively high at the beginning of the course. We found that students' perceived utility value changed during the course of the school term, with a significant decline from the beginning to the middle of the course, regardless of gender or race. This finding is consistent with prior findings among secondary and undergraduate students (Gaspard et al., 2015; Jacobs et al., 2002; Robinson et al., 2019) as well as research that has examined short-term changes in motivation (Kosovich, Flake, et al., 2017; Perez et al., 2014).

6.2. Utility value and changes in utility value across demographic subgroups

No significant differences in utility value were observed by gender, indicating that female and male students entered the introductory statistics course with similar levels of perceived utility value and remained on similar levels throughout the term. Similarly, we did not find evidence of gender differences in students' perceived utility value change over the term: Female and male students' utility value developed similarly over the term, which is in line with prior research (Robinson et al., 2019). However, there were significant differences in behavioral engagement at the beginning of the course and in final grades by gender, with male students having significantly higher behavioral engagement scores and final course grades than female students.

The story was different based on students' URM membership: Although we found no differences in average levels of utility value at the beginning or mid-course, there were significant differences by the end of the course with Black, Latinx and other underrepresented minoritized students reporting significantly lower levels of utility value than White/Asian students. These findings suggest a growing disparity in utility value of the course. URM students also had significantly lower behavioral engagement scores at all three time points and received lower final grades. In line with this trend, URM students continued to experience a decline in values toward the end of the course. Latent change analyses revealed differential rates of change in utility value with significant differences from the middle to the end of the course, with the change being significantly more negative for URM students. Consistent with a large body of research on differences in academic performance between minoritized students and majority students (Board, 2018), URM students in this study had significantly lower course grades than non-URM students. There seems to be ongoing equity gaps between groups based on race/ethnicity, which may be attributed to systemic structures and processes that likely instigate opportunity gaps (e.g., differences in the opportunity to learn, resources for education, quality of teachers, socioeconomic factors), achievement gaps (i.e., grades, test scores), and psychological factors not reported in this study such as their success expectancies, perceived costs associated with the course, sense of belonging in college, or stereotype threat (Urdan & Herr, 2017). Although URM students are more likely to be the first member of one's family to attend college and to come from economically disadvantaged families, it is important to remember that URM students are a very heterogeneous group whose members differ in generational and socioeconomic status as well country of origin and first language, which in turn affects their academic motivation (Urdan & Herr, 2017). For example, among Latinx students, second-generation students (i.e., children whose parents immigrated to the United States) have been found to have higher academic aspirations and more positive motivational profiles than third generation students (i.e., children of parents born in the United States) (Schleicher, 2006; Urdan & Herr, 2017). Thus, further research capturing the intersectionality between different racial backgrounds and generation status is warranted to better understand.

The present study adds to prior research by examining change in utility value by gender and race and highlights the ongoing concern for underrepresented groups in the sciences. The differential trajectories in utility value highlight the necessity to grasp how institutions are serving students differently, which may produce disparities in motivation and motivational profiles in the context of statistics education. Because we only measured utility value in our study, future research should simultaneously examine the trajectories of different motivational variables (e.g., value *and* expectancy levels) to determine how "one construct [that] fluctuates during a semester may be related or unrelated to how the other fluctuates" (Kosovich, Flake, et al., 2017, p. 132), while accounting for gender, race, and racial minority status. Further, because different types of values (i.e., intrinsic, utility, and attainment value) have been suggested to differentially predict learning outcomes (Eccles, 2009), examining change trajectories in said different facets of values would greatly inform education and intervention practice (Kosovich, Flake, et al., 2017).

6.3. Relationship between utility value, behavioral engagement in the learning process, and course grades

There were reciprocal relationships between utility value and behavioral engagement in the course, with behavioral engagement during the course predicting their utility value at the end of the course and vice versa. Specifically, perceived utility value of the course during the course was predictive of their behavioral engagement at the end of the course. Both utility value and behavioral engagement positively predicted students' final course grade. Although value perceptions have been found to be directly and positively linked to performance (e.g., Wigfield & Eccles, 2002), several studies have shown that they are stronger predictors of future intentions and choice-related behaviors (e.g., enrolling in STEM courses) and continued interest, whereas success expectancies within the expectancy-value framework are stronger predictors of performance (Acee & Weinstein, 2010; Wigfield & Eccles, 2000). Thus, future research should simultaneously examine course related expectations and values as well as include choice-related behaviors such as students' intentions and interests to enroll in a statistics course in the future.

Overall, this pattern of results is compelling with regard to the malleability of utility value as well as implications for motivational interventions because they suggest that whereas students' incoming, preexisting utility value perceptions about statistics may not serve as a predictor of course outcomes, they can change as a result of students experiences in educational environments (Robinson et al., 2019). In other words, while instructors typically do not have the ability to influence students' incoming utility value, they - along with motivationally supportive instructional material - could impact how students' utility value changes throughout the term (Young et al., 2018). Over the past decade, there has been a growing body of research focused on improving student learning and learning outcomes, especially in STEM subjects, by implementing expectancy-value based interventions (Gaspard et al., 2015; Kosovich et al., 2019; Rosenzweig et al., 2020; Tibbetts et al., 2016). Such interventions are designed to facilitate perceptions of utility value through emphasis on the relevance of the course topic or coursework for students' lives. When students believe what they are learning in a course is useful, relevant, and applicable to their lives, they tend to be more interested in the course topic, become more engaged in the material, and are more successful in class (Hulleman et al., 2008; Kosovich et al., 2019). Not only have utility-value interventions been shown to be effective across different academic domains, including biology (Brown et al., 2015; Canning et al., 2018; Harackiewicz et al., 2014), math (Gaspard et al., 2015; Kosovich et al., 2019), and statistics (Acee & Weinstein, 2010), but also across school levels, including elementary (Shin et al., 2019), secondary (Gaspard et al., 2015; Hulleman & Harackiewicz, 2009), and college (Canning et al., 2018; Hulleman et al., 2017).

These interventions have been particularly effective in supporting the motivation and learning outcomes for URM students, including first-generation, Black and Latinx, and students from lower socioeconomic backgrounds (e.g., Brown et al., 2015; Harackiewicz et al., 2014; Tibbetts et al., 2016). Our results suggest that all students would benefit from a more motivationally supportive learning context in the first half of the semester, and that Black and Latinx students would benefit from more motivational supports in the second half of the semester. In addition to providing the opportunity to participate in utility-value interventions, curricular materials and course activities could be embedded with messages related to prosocial and altruistic goals (i.e., demonstrating how data science and statistics topics can be used to make a positive impact on the world), which are messages that often resonate with students from racially minoritized backgrounds (Jackson et al., 2016; Thoman et al., 2015, 2017), or emphasize the role of collaboration in the learning and practice of statistics (Allen et al., 2015).

7. Limitations and avenues for future research

Some important limitations must be considered when interpreting the findings of this study. First, the data for this study came from one institution and one introductory statistics course, thus, it is unclear whether the same trajectories would replicate among students from other institutions and in other statistics intro course materials. Given that the acceptance rate at the institution within which this study was conducted is relatively low (around 12%), the results may generalize only to students at selective institutions. Future research should be conducted within different institutional contexts.

Second, utility value was measured using only two items, which can undermine reliability and validity. Although research exists that only used two items to measure utility value (Kosovich et al., 2019) and even suggest that single-item measures can be appropriate for unidimensional constructs (e.g., Gogol et al., 2014), it must be acknowledged as a limitation that limits the interpretability of the findings.

Third, the content of the course in which data were collected increased in difficulty over time. The current analyses do not take into consideration how changes in difficulty might have interacted with changes in behavioral engagement and performance. Perhaps students who struggled with the content used utility value as a scapegoat, reducing their ratings of value to protect their self-image. Or, perhaps there was a level of difficulty that, once reached, caused students to disengage with the material. Future research should investigate these complex interactions (or be conducted in a course with a stable level of difficulty throughout).

Fourth, the indicator of behavioral engagement was measured using the scores from the review question at the end of each chapter, which was chosen to reflect students' participation and the quality of their engagement in the learning process. Although there is "no single correct definition of *engagement*" (Skinner et al., 2009, p. 224), educational research typically considers self-reported engagement measures. While we consider the combination of self-report measure of utility value and direct measure of engagement and performance a strength, we acknowledge that it not only limits comparability to prior research, but that there is also the possibility that the engagement in the review questions varied depending on attendance rates with those who have poor course attendance potentially using this learning activity to compensate for their absences. Nevertheless, the fact that we did find the expected correlations based on expectancy-value frameworks (i.e., between behavioral engagement and utility value as well as between behavioral engagement and grades), suggests that the measure of behavioral engagement shows some validity in this study.

Finally, some small effect sizes and path coefficients (e.g., 0.06; −0.08) should be interpreted with caution.

In addition to the avenues for future research mentioned previously, an important future direction lies in gaining a deeper understanding of students' psychological experiences - in particular among URM and Latinx students - in introductory statistics in order to (a) identify what will work to motivate students to learn, persist, and succeed in introductory statistics courses as well as (b) understand how to best support those struggling most. Based on the current findings, an important future direction relates to changes in the course that could be made in order to better support the motivation of all students, and for URM students in particular. The overall trajectory of utility value suggests that interventions during the first half of the course may be most effective in order to mitigate the rate of decline in motivation. This is supported by the fact that initial utility value does not predict learning outcomes directly; instead, it's the value that students see in the course in the middle and end of the semester that are related to learning outcomes. This could be particularly true for minoritized students who are less likely to see value in STEM courses for a variety of reasons, including a lack of models and perceived cultural mismatches (Tibbetts et al., 2016). These barriers can be partially overcome by direct-to-student interventions that scaffold students to see how STEM courses can be relevant to their lives and help them achieve prosocial and communal goals (Allen et al., 2015; Brown et al., 2015; Jackson et al., 2016; Thoman et al., 2015). Additionally, given the interplay between utility value and behavioral engagement, interventions should be designed in a comprehensive manner that addresses both the relevance of the content as well as the importance of engaging in learning material to maximize students' course success.

A unique feature of this study is that the introductory statistics course was based on an online interactive textbook, providing another interesting avenue for future research. Although the aim of this study was not to compare different course formats, but instead to explore the role of utility value and behavioral engagement in this particular context, exploring teaching and learning using online material is critical not only given its rapid growth in colleges and universities across the globe (Songsore & White, 2018), but also because research has pointed to the use of technology to engage students in learning statistical concepts, which in turn leads to higher retainment of the content (Marchionda & Autin, 2016).

Finally, future research should explore instructor behavior and how it might influence perceptions of utility value. While messages about the usefulness of the course content and material were more implicit than explicit (e.g., students analyzed datasets from actual psychology studies), how instructors discuss the utility of using R in lectures and discussions are likely to vary.

8. Conclusions

Although a large body of research indicates that value beliefs and similar motivational constructs tend to decline over the long-term and over key transition points (Jacobs et al., 2002; Musu-Gillette et al., 2015), the present study adds to the sparse research exploring short-term changes in utility value trajectories in introductory college statistics, accounting for gender and underrepresented racial minority status. Research on motivation in higher education is a key effort to deepen our understanding of what will work to motivate college students to learn, persist, and succeed in introductory statistics courses as well as to understand how to best support those struggling most. The findings highlight that although students show relatively high levels of initial utility value, students' perceptions of the usefulness and importance of the course declined from the beginning to the middle of the course. Not only did we observe differential rates of decline, we further found differential trajectories from the middle to the end of the course by underrepresented racial minority status. Moreover, students' perceived utility value was reciprocally related to students' behavioral engagement during the learning process, and both predicted final course grade.

Funding

This project has been made possible in part by a grant from the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation (DRL-1229004), a grant from the California Governor's Office of Planning and Research (OPR18115), and a grant from the National Science Foundation (EHR-2000507).

Declaration of competing interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.lindif.2021.102095.

References

Acee, T. W., & Weinstein, C. E. (2010). Effects of a value-reappraisal intervention on statistics students' motivation and performance. *The Journal of Experimental Education, 78*(4), 487–512. https://doi.org/10.1080/00220970903352753

Allen, J., Muragishi, G. A., Smith, J. L., Thoman, D. B., & Brown, E. R. (2015). To grab and to hold: Cultivating communal goals to overcome cultural and structural barriers in first generation college students' science interest. *Translational Issues in Psychological Science, 1*, 331–341. https://doi.org/10.1037/tps0000046

Armenta, B. E. (2010). Stereotype boost and stereotype threat effects: The moderating role of ethnic identification. *Cultural Diversity and Ethnic Minority Psychology, 16*(1), 94–98. https://doi.org/10.1037/a0017564

Barron, K. E., & Hulleman, C. S. (2015). xpectancy-value-cost model of motivation. In J. S. Eccles, & K. Salmelo-Aro (Eds.) (2nd ed.,*Motivational psychologyInternational Encyclopedia of social and behavioral sciences* (pp. 503–509). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.26099-6.

Bateiha, S., Marchionda, H., & Autin, M. (2020). Teaching style and attitudes: A comparison of two collegiate introductory statistics classes. *Journal of Statistics Education, 28*(2), 154–164. https://doi.org/10.1080/10691898.2020.1765710

Bathgate, M., & Schunn, C. (2017). Factors that deepen or attenuate decline of science utility value during the middle school years. *Contemporary Educational Psychology, 49*, 215–225. https://doi.org/10.1016/j.cedpsych.2017.02.005

Board, N. S. (2018). *Science and engineering indicators 2018*. Arlington, VA: National Science Foundation.

Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociological Methodology, 37–69*. https://doi.org/10.2307/271028

Brown, E. R., Smith, J. L., Thoman, D. B., Allen, J. M., & Muragishi, G. (2015). From bench to bedside: A communal utility value intervention to enhance students' biomedical science motivation. *Journal of Educational Psychology, 107*(4), 1116–1135. https://doi.org/10.1037/edu0000033

Canning, E. A., Harackiewicz, J. M., Priniski, S. J., Hecht, C. A., Tibbetts, Y., & Hyde, J. S. (2018). Improving performance and retention in introductory biology with a utility-value intervention. *Journal of Educational Psychology, 110*(6), 834. https://doi.org/10.1037/edu0000244

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modelling, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Chiesi, F., & Primi, C. (2010). Cognitive and non-cognitive factors related to students' statistics achievement. Online *Statistics Education Research Journal, 9*(1), 6–26 http://www.stat.auckland.ac.nz/~iase/serj/SERJ9%281%29_Chiesi_Primi.pdf.

Chouinard, R., & Roy, N. (2008). Changes in high-school students' competence beliefs, utility value and achievement goals in mathematics. *British Journal of Educational Psychology, 78*(1), 31–50. https://doi.org/10.1348/000709907X197993

Cohen, J., West, S. G., Aiken, L. S., & Cohen, P. (2002). *Applied multiple regression/correlation analysis for behavioral sciences* (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Crede, M., & Phillips, L. A. (2011). A meta-analytic review of the motivated strategies for learning questionnaire. *Learning and Individual Differences, 21*, 337–346. https://doi.org/10.1016/j.lindif.2011.03.002

Durik, A. M., Vida, M., & Eccles, J. S. (2006). Task values and ability beliefs as predictors of high school literacy choices: A developmental analysis. *Journal of Educational Psychology, 98*(2), 382–393. https://doi.org/10.1037/0022-0663.98.2.382

Eccles, J. S. (2005). Subjective task values and the model of achievement related choices. In A. J. Elliott, & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 105–121). New York, NY: Guilford.

Eccles, J. S. (2007). Where are all the women? Gender differences in participation in physical science and engineering. In S. J. Ceci, & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 199–210). Washington, DC: American Psychological Association. https://doi.org/10.1037/11546-016.

Eccles, J. S. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist, 44*(2), 78–89. https://doi.org/10.1080/00461520902832368

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146). San Francisco: W. H. Freeman.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Fong, C. J., Kremer, K. P., Troglin-Cox, C. H., & Lawson, C. A. (2021). Expectancy-value profiles in math and science: A person-centered approach to cross-domain motivation with academic and STEM-related outcomes. *Contemporary Educational Psychology*. https://doi.org/10.1016/j.cedpsych.2021.101962

Fredricks, J. A., Bohnert, A. M., & Burdette, K. (2014). Moving beyond attendance: Lessons learned from assessing engagement in afterschool contexts. *New Directions for Youth Development, 2014*(144), 45–58. https://doi.org/10.1002/yd.20112

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology, 38*(4), 519–533. https://doi.org/10.1037/0012-1649.38.4.519

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In *Handbook of research on student engagement* (pp. 763–782). Boston, MA: Springer.

Friedrich, J., Buday, E., & Kerr, D. (2000). Statistical training in psychology: A national survey and commentary on undergraduate programs. *Teaching of Psychology, 27*(4), 248–257. https://doi.org/10.1207/S15328023TOP2704_02

Friedrich, J., Childress, J., & Cheng, D. (2018). Replicating a national survey on statistical training in undergraduate psychology programs: Are there "new statistics" in the new millennium? *Teaching of Psychology, 45*(4), 312–323. https://doi.org/10.1177/0098628318796414

Gaspard, H., Dicke, A.-L., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (2015). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology, 51*(9), 1226–1240. https://doi.org/10.1037/dev0000028

Geiser, C. (2012). *Data analysis with Mplus*. Chicago: Guilford press. https://doi.org/10.1007/978-3-531-93192-0

Gibbens, B. (2019). Measuring student motivation in an introductory biology class. *The American Biology Teacher, 81*(1), 20–26.

Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., Fischbach, A., & Preckel, F. (2014). "My questionnaire is too Long!" The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology, 39*, 188–205. https://doi.org/10.1016/j.cedpsych.2014.04.002

Goudas, A. M., & Boylan, H. R. (2012). Addressing flawed research in developmental education. *Journal of Developmental Education, 36*(1), 2–4, 6,8,10,12–13.

Gray, D. L., Hope, E. C., & Matthews, J. S. (2018). Black and belonging at school: A case for interpersonal, instructional, and institutional opportunity structures. *Educational Psychologist, 53*(2), 97–113. https://doi.org/10.1080/00461520.2017.1421466

Guo, J., Parker, P. D., Marsh, H. W., & Morin, A. J. S. (2015). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. *Developmental Psychology, 51*(8), 1163–1176. https://doi.org/10.1037/a0039440

Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Giffen, C. J., Blair, S. S., Rouse, D. I., & Hyde, J. S. (2014). Closing the social class achievement gap for first-generation students in undergraduate biology. *Journal of Educational Psychology, 106*(2), 375–389. https://doi.org/10.1037/a003467

Harris, R. B., Mack, M. R., Bryant, J., Theobald, E. J., & Freeman, S. (2020). *ScienceAdvances, 6*(24). https://doi.org/10.1126/sciadv.aaz5687

Hood, M., Creed, P. A., & Neumann, D. L. (2012). Using the expectancy value model of motivation to understand the relationship between student attitudes and achievement in statistics. *Statistics Education Research Journal, 11*(2), 72–85.

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

Hulleman, C. S. (2007). *The role of utility value in the development of interest and achievement*. University of Wisconsin-Madison. Unpublished doctoral dissertation.

Hulleman, C. S., Durik, A. M., Schweigert, S. B., & Harackiewicz, J. M. (2008). Task values, achievement goals, and interest: An integrative analysis. *Journal of Educational Psychology, 100*(2), 398–416. https://doi.org/10.1037/0022-0663.100.2.398

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology, 102*(4), 880–895. https://doi.org/10.1037/a0019506

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science, 326*(5958), 1410–1412. https://doi.org/10.1126/science.1177067

Hulleman, C. S., Kosovich, J. J., Barron, K. E., & Daniel, D. B. (2017). Making connections: Replicating and extending the utility value intervention in the classroom. *Journal of Educational Psychology, 109*(3), 387–404. https://doi.org/10.1037/edu0000146

Jackson, M. C., Galvez, G., Landa, I., Buonora, P., & Thoman, D. B. (2016). Science that matters: The importance of a cultural connection in underrepresented students' science pursuit. *CBE Life Sciences Education*. https://doi.org/10.1187/cbe.16-01-0067

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73*(2), 509–527. https://doi.org/10.1111/1467-8624.00421

Kosovich, J. J., Flake, J. K., & Hulleman, C. S. (2017). Understanding short-term motivation trajectories: A parallel process model of expectancy-value motivation. *Contemporary Educational Psychology, 49*, 130–139. https://doi.org/10.1016/j.cedpsych.2017.01.004

Kosovich, J. J., Hulleman, C. S., & Barron, K. E. (2017). Measuring motivation in educational settings: A case for pragmatic measurement. To appear in. In K. A. Renninger, & S. E. Hidi (Eds.), *The Cambridge handbook on motivation and learning* (pp. 39–60). New York, NY: Routledge.

Kosovich, J. J., Hulleman, C. S., Barron, K. E., & Getty, S. (2015). A practical measure of student motivation: Establishing validity evidence for the expectancy-value-cost scale in middle school. *The Journal of Early Adolescence, 35*(5–6), 790–816. https://doi.org/10.1177/0272431614556890

Kosovich, J. J., Hulleman, C. S., Phelps, J., & Lee, M. (2019). Improving algebra success with a utility-value intervention. *Journal of Developmental Education, 42*(2), 2–10.

Krumm, A. E., Beattie, R., Takahashi, S., D'Angelo, C., Feng, M., & Cheng, B. (2016). Practical measurement and productive persistence: Strategies for using digital learning system data to drive improvement. *Journal of Learning Analytics, 3*(2), 116–138. https://doi.org/10.18608/jla.2016.32.6

Lee, V. E. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist, 32*, 125–141. https://doi.org/10.1207/S15326985EP3502_6

Little, T. D. (2013). *Longitudinal structural equation modelling*. Guilford Press.

Marchionda, H., & Autin, M. (2016). Using technology to engage students in introductory statistics. Retrieved from. In , *Shining a light on mathematics learningShining a light on mathematics learning* (pp. 158–165) https://www.rcml-math.org/assets/Proceedings/rcml_proceedings–2016.pdf#page=164.

Martin, A. J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *British Journal of Educational Psychology, 77*(2), 413–440. https://doi.org/10.1348/000709906X118036

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543. https://doi.org/10.1007/BF02294825

Musu-Gillette, L. E., Wigfield, A., Harring, J. R., & Eccles, J. S. (2015). Trajectories of change in students' self-concepts of ability and values in math and college major choice. *Educational Research and Evaluation, 21*(4), 343–370. https://doi.org/10.1080/13803611.2015.1057161

Muthén, L., & Muthén, B. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Newsom, J. T. (2015). *Longitudinal structural equation modelling: A comprehensive introduction*. New York, NY: Routledge. https://doi.org/10.4324/9781315871318

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

Ramirez, C., Emmioğlu, E., & Schau, C. (2010, August). Understanding students' attitudes toward statistics: New perspectives using an expectancy-value model of motivation and the survey of attitudes toward statistics. In *Paper presented at joint statistical meetings, Vancouver*. http://www.evaluationandstatistics.com/JSM2010EVM.pdf.

Rhoads, T. R., & Hubele, N. F. (2000). Student attitudes toward statistics do. *IEEE Transactions on Education, 43*(2), 182–187. https://doi.org/10.1109/13.848071

Riegle-Crumb, C., King, B., & Irizarry, Y. (2019). Does STEM stand out? Examining racial/ethnic gaps in persistence across postsecondary fields. *Educational Researcher, 48*(3), 133–144. https://doi.org/10.3102/0013189X19831006

Robinson, K. A., Lee, Y.-K., Bovee, E. A., Perez, T., Walton, S. P., Briedis, D., & Linnenbrink-Garcia, L. (2019). Motivation in transition: Development and roles of

expectancy, task values, and costs in early college engineering. *Journal of Educational Psychology, 111*(6), 1081–1102. https://doi.org/10.1037/edu0000331
Rosenzweig, E. Q., Hulleman, C. S., Barron, K. E., Kosovich, J. J., Priniski, S. J., & Wigfield, A. (2019). Promises and pitfalls of adapting utility value interventions for online math courses. *The Journal of Experimental Education, 87*(2), 332–352. https://doi.org/10.1080/00220973.2018.1496059
Rosenzweig, E. Q., & Wigfield, A. (2016). STEM motivation interventions for adolescents: A promising start, but further to go. *Educational Psychologist, 51*(2), 146–163. https://doi.org/10.1080/00461520.2016.1154792
Rosenzweig, E. Q., Wigfield, A., & Hulleman, C. S. (2020). More useful or not so bad? Examining the effects of utility value and cost reduction interventions in college physics. *Journal of Educational Psychology, 112*(1), 166–182. https://doi.org/10.1037/edu0000370
Rybczynski, S. M., & Schussler, E. E. (2013). Effects of instructional model on student attitude in an introductory biology laboratory. *International Journal for the Scholarship of Teaching and Learning, 7*(2), Article n2. https://doi.org/10.20429/ijsotl.2013.070222
Schau, C., & Emmioğlu, E. (2012). Do introductory statistics courses in the United States improve students' attitudes? *Statistics Education Research Journal, 11*(2), 86–94.
Schau, C., Stevens, J. J., Dauphinee, T. L., & Del Vecchio, A. M. (1995). The development and validation of the survey of attitudes toward statistics. *Educational and Psychological Measurement, 55*(5), 868–875. https://doi.org/10.1177/0013164495055005022
Schleicher, A. (2006). Where immigrant students succeed: A comparative review of performance and engagement in PISA 2003. *Intercultural Education, 17*(5), 507–516. https://doi.org/10.1080/14675980601063900
Shin, D. D., Lee, M., Ha, J. E., Park, J. H., Ahn, H. S., Son, E., Chung, Y., & Bong, M. (2019). Science for all: Boosting the science motivation of elementary school students with utility value intervention. *Learning and Instruction, 60*, 104–116. https://doi.org/10.1016/j.learninstruc.2018.12.003
Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology, 100*(4), 765–781. https://doi.org/10.1037/a0012840
Skinner, E. A., Kindermann, T. A., Connell, J. P., & Wellborn, J. G. (2009). Engagement and disaffection as organizational constructs in the dynamics of motivational development. In K. R. Wenzel, & A. Wigfield (Eds.), *Educational psychology handbook series. Handbook of motivation at school* (pp. 223–245). Routledge/Taylor & Francis Group.
Songsore, E., & White, B. J. G. (2018). Students' perceptions of the future relevance of statistics after completing an online introductory statistics course. *Statistics Education Research Journal, 17*(2), 120–140. http://www.stat.auckland.ac.nz/serj.
Sorge, C., & Schau, C. (2002, April). Impact of engineering students' attitudes on achievement in statistics. In *Paper presented at the annual meeting of the American Educational ResearchAssociation, New Orleans*. http://evaluationandstatistics.com/AERA2002.pdf.
Strayhorn, T. L. (2012). *College students' sense of belonging: A key to educational success for all students.* New York, NY: Routledge.
Thoman, D. B., *Muragishi, G. A., & Smith, J. L. (2017). Research microcultures as socialization contexts for underrepresented science students. *Psychological Science*. https://doi.org/10.1177/0956797617694865. Advanced Online Publication.
Thoman, D. B., Brown, E. R., Mason, A. Z., Harmsen, A. G., & Smith, J. L. (2015). The role of altruistic values in motivating underrepresented minority students for biomedicine. *BioScience, 65*, 183–188. https://doi.org/10.1093/biosci/biu19
Tibbetts, Y., Harackiewicz, J. M., Priniski, S. J., & Canning, E. A. (2016). Broadening participation in the life sciences with social–psychological interventions. *CBE—Life sciencesEducation, 15*(3), es4. https://doi.org/10.1187/cbe.16-01-0001
Urdan, T., & Herr, V. (2017). Motivation and achievement of Hispanic college students in the United States. In J. T. DeCuir-Gunby, & P. A. Schutz (Eds.), *Race and ethnicity in the study of motivation in education* (pp. 178–192). NY: Routledge. https://doi.org/10.4324/9781315716909-18.
Van Es, C., & Weaver, M. M. (2018). Race, sex, and their influences on introductory statistics education. *Journal of Statistics Education, 26*(1), 48–54. https://doi.org/10.1080/10691898.2018.1434426
Wang, M. T., & Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy-value perspective to understand individual and gender differences in STEM fields. *Developmental Review, 33*(4), 304–340. https://doi.org/10.1016/j.dr.2013.08.001
Watt, H. M. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th-through 11th-grade Australian students. *Child Development, 75*(5), 1556–1574. https://doi.org/10.1111/j.1467-8624.2004.00757.x
Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance abuse domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). https://doi.org/10.1037/10222-009
Wiebe, E., Unfried, A., & Faber, M. (2018). The relationship of STEM attitudes and career interest. *EURASIA Journal of Mathematics, Science and Technology Education, 14*(10). https://doi.org/10.29333/ejmste/92286
Wigfield, A., & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review, 30*(1), 1–35. https://doi.org/10.1016/j.dr.2009.12.001
Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81. https://doi.org/10.1016/j.dr.2009.12.001
Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield, & J. S. Eccles (Eds.), *A Vol. in the Educational Psychology Series. Development of Achievement Motivation* (pp. 91–120). Academic Press. https://doi.org/10.1016/B978-012750053-9/50006-1.
Wigfield, A., Eccles, J. S., Fredricks, J. A., Simpkins, S., Roeser, R. W., & Schiefele, U. (2015). Development of achievement motivation and engagement. In M. E. Lamb, & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science: Socioemotional processes* (pp. 657–700). John Wiley & Sons, Inc.. https://doi.org/10.1002/9781118963418.childpsy316
Wigfield, A., Tonks, S., & Klauda, S. T. (2009). Expectancy-value theory. In A. Wentzel, K. R.W. (Ed.), *Handbook of motivation at school* (pp. 55–75). New York, NY: Routledge.
Young, A. M., Wendel, P. J., Esson, J. M., & Plank, K. M. (2018). Motivational decline and recovery in higher education STEM courses. *International Journal of Science Education, 40*(9), 1016–1033. https://doi.org/10.1080/09500693.2018.1460773
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. https://doi.org/10.1080/0950069032000052207