A Framework of College Student Buy-in to Evidence-Based Teaching Practices in STEM: The Roles of Trust and Growth Mindset

**Cong Wang,[†] Andrew J. Cavanagh,[‡] Melanie Bauer,[†] Philip M. Reeves,[§]
Julia C. Gill,[†] Xinnian Chen,[‖] David I. Hanauer,[¶] and Mark J. Graham[†]***

[†]STEM Program Evaluation & Research Lab (STEM-PERL), Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06420; [‡]Educational Alliance, New York, NY 10002; [§]School of Medicine, Johns Hopkins University, Baltimore, MD 21205; [‖]Department of Physiology and Neurobiology, University of Connecticut, Storrs, CT 06269; [¶]Department of English, Indiana University of Pennsylvania, Indiana, PA 15705

ABSTRACT

Evidence-based teaching practices (EBPs) foster college science, technology, engineering, and mathematics (STEM) students' engagement and performance, yet our knowledge of what contributes to the effectiveness of these practices is less established. We propose a framework that links four social-cognitive variables—students' trust in their instructors, growth mindset, buy-in to instructional practices, and course engagement—to long-standing desired student outcomes of academic performance and intent to persist in science. This framework was tested in classrooms identified as having a high level of EBP implementation with a multi-institutional sample of 2102 undergraduates taught by 14 faculty members. Results indicate that the buy-in framework is a valid representation of college students' learning experiences within EBP contexts overall as well as across underrepresented student groups. In comparison to students' level of growth mindset, students' trust in their instructors was more than twice as predictive of buy-in to how the course was being taught, suggesting that students' views of their instructors are more associated with thriving in a high-EBP course environment than their views of intelligence. This study contributes to the dialogue on transforming undergraduate STEM education by providing a validated student buy-in framework as a lens to understand how EBPs enhance student outcomes.

INTRODUCTION

There is variability in the degree to which students accept evidence-based teaching practices (EBPs) and learn from them (Brazeal *et al.*, 2016; Nguyen *et al.*, 2017; Bonem *et al.*, 2020; Finkelstein *et al.*, 2019; Starr *et al.*, 2020). EBPs are research-supported activities that promote student engagement and learning through active learning, formative assessment, metacognition, and inclusive teaching practices (Handelsman *et al.*, 2007; Graham *et al.*, 2013; Starr *et al.*, 2020; Theobald *et al.*, 2020). Research on EBPs shows benefit for college science, technology, engineering, and mathematics (STEM) students (Freeman *et al.*, 2014), and particularly for members of traditionally underrepresented groups in the sciences (Gross *et al.*, 2015; Theobald *et al.*, 2020). Nevertheless, other research indicates that some students show resistance to learning in ways prescribed by EBPs and therefore miss out on the intended benefits (Patrick, 2020).

Despite national calls for undergraduate STEM education reform (Woodin *et al.*, 2010), there has yet to be widespread implementation of EBPs by instructors, and the lecture remains commonplace (Stains *et al.*, 2018). Moving the field forward requires

further investigation into the variability in college science students' commitment to EBPs (Corwin *et al.*, 2015). We propose that a deeper understanding of the social and cognitive processes that underlie students' buy-in to the EBP learning process will enable instructors to be more deliberate in fostering student acceptance of EBPs, improve students' feelings about these practices, and expand instructors' use of them.

The level of college student buy-in to active learning and other EBPs in a science course, along with its ability to increase desired student outcomes, was initially explored by Cavanagh *et al.* (2016) based upon a faculty buy-in framework for adopting EBPs in their teaching (Aragón *et al.*, 2017). The research was conducted in one large introductory science course taught by one instructor. This preliminary work had students make judgments about the extent to which they believed that various EBPs that they had been exposed to, such as group work, were: 1) educationally valuable and 2) beneficial to their own learning. The resulting framework of exposure–persuasion–identification–commitment (EPIC; Cavanagh *et al.*, 2016) provided a method for describing and measuring different levels of a student's buy-in to an instructor's use of EBPs. It also demonstrated significant positive associations between student commitment (i.e., buy-in), engagement in self-regulated learning, and subsequent course performance.

In addition to testing the effects of student buy-in itself, factors that influence buy-in are also important. Drawing from an interpersonal relationship theory (Reis and Gable, 2015) and a growth mindset perspective (Dweck, 2008), in a follow-up study Cavanagh *et al.* (2018) tested the associations of two social-cognitive variables—trust and growth mindset—with student buy-in to EBPs. This work confirmed that both trust and growth mindset were significant predictors of student buy-in.

The justification for testing students' growth mindset as part of the buy-in framework is well grounded in the psychology and education literatures as a consistent predictor of educational success for adolescents and young adults (Dweck, 2008; Yeager *et al.*, 2016; Degol *et al.*, 2018). Students who possess a growth mindset tend to view learning as an opportunity to expand their knowledge and intelligence. In contrast, a fixed mindset represents a view that intelligence is a static entity that is relatively resistant to change and experience. Therefore, students with a growth mindset are more likely to view classroom experiences as learning opportunities and embrace EBPs (Cavanagh *et al.*, 2018).

In contrast to growth mindset, trust, as studied in the college classroom, is a relatively new construct that may also relate to facilitating student buy-in within active-learning contexts. In terms of background, it has been well established by a variety of studies that interaction between instructors and students plays a critical role in student learning in undergraduate contexts (Komarraju *et al.*, 2010; Kim and Sax, 2014; Fedesco *et al.*, 2019; Snijders *et al.*, 2020). For instance, Fedesco *et al.* (2019) found that college students who felt more connected to their instructors demonstrated more interest in a course, put forth more effort, and perceived they learned more. Students' trust in their instructors refers to "a perception that the instructor understands the challenges facing students as they progress through the course, accepts students for who they are, and cares about the educational welfare of students" (Cavanagh *et al.*, 2018, p. 2). The extent to which students trust their instructors—that is, view them as understanding, accepting, and caring—is likely to predict their attitudes toward courses and willingness to engage in course activities. This concept of trust is particularly important to examine in college STEM courses, as many disciplines engage instructors who are highly regarded researchers but often undervalue the importance of instructor–student relationships (Christe, 2013).

Although previous work has shed some light on the social-cognitive factors influencing college students' engagement and performance (Cavanagh *et al.*, 2016, 2018), it has at least two important limitations. First, previous work was conducted in only one large college science course with one instructor (Patrick, 2020). Therefore, there is a need for testing the effects of student buy-in on a broader sample of instructors and students across multiple contexts. Second, a more integrated analysis of the social-cognitive factors is needed. Cavanagh and colleagues' analysis was limited in that it independently tested the contributions of growth mindset and trust to buy-in, engagement, and course grades rather than testing a simultaneous, fully linked buy-in framework (Cavanagh *et al.*, 2018). The current study addresses these two limitations by using a multi-institutional sample to test a comprehensive student buy-in framework.

THE CURRENT STUDY

The current investigation focused on two key research questions. Our primary question concerned whether the hypothesized buy-in framework (see Figure 1), which is an extension of prior research from Cavanagh et al. (2016, 2018), can be used to understand college students' learning experiences across instructors in STEM classrooms featuring EBPs. Based on the literature, we hypothesize that students' trust in instructors and growth mindset will enhance college students' buy-in toward EBPs, through the pathway from persuasion to identification to commitment. In turn, this will foster student engagement and subsequently affect desired student outcomes, including intent to persist in science and final course grade. To understand the generalizability of the proposed buy-in framework, as a follow-up analysis, we further examined whether the variable interrelationships hold in the same way across different racial/ethnic groups and between male and female students.

The present study contributes to the existing literature on college STEM education reform in several ways. It is the first study to evaluate the tenets of a buy-in framework that links students' trust, growth mindset, buy-in, and engagement with intent to persist in science and academic performance. Second, the present study was conducted across a number of science courses and institutions with a large student sample. Compared with the other studies with student samples within a large single course (Cavanagh *et al.*, 2016, 2018), the findings of the current study answer calls to be more generalizable in terms of researching broader contexts (Patrick, 2020) and therefore are more informative for educational practice and policy. Third, it is more statistically robust because it uses the technique of structural equation modeling (SEM) with latent variables, through which we were able to explicitly specify the measurement errors associated with the constructs, obtain better estimates of the true interrelationships among the variables of interest, and simultaneously test multiple relationships specified in the buy-in framework.
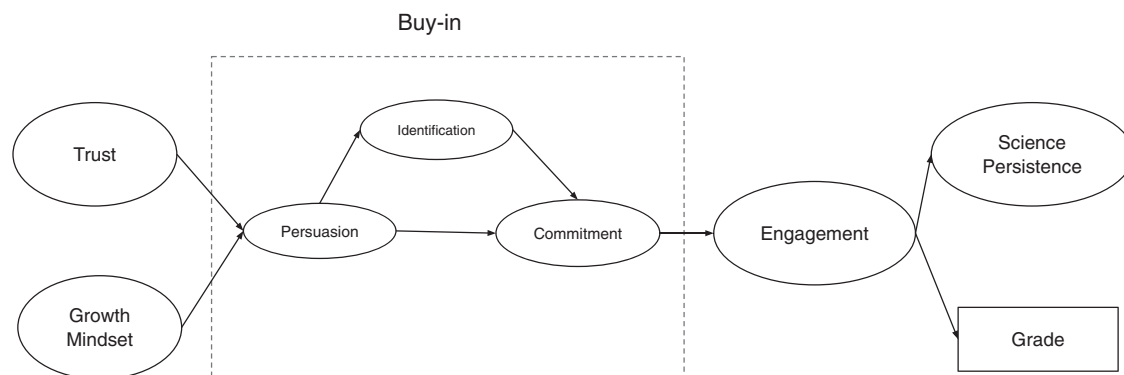
FIGURE 1. The proposed buy-in framework of college students' learning experiences in transformed STEM classes. Rectangles represent observed variables and ovals represent latent variables in the path diagram.

METHODS

Participants

The current study's sampling procedure consisted of two phases. First, we identified a sample of instructors who had attended at least one Summer Institute on Scientific Teaching (SI) and who were implementing a high number of EBPs. Second, we surveyed students in courses taught by these instructors.

In terms of background, the SI is a four-and-a-half-day intensive faculty development training focused on active learning, formative assessment, and inclusive teaching practices. The SI was developed for college STEM faculty and instructors (Handelsman *et al.*, 2007). During the SI, participants engage in sessions on inclusive teaching, active learning, assessment, how people learn, and institutional transformation. Participants also work in teams to develop a 15-minute teachable unit incorporating research-based instructional strategies. The SI's larger aim is to transform undergraduate science courses to be more active, engaging, and evidence based. For more description of the SI and faculty adoption of its principles in the college STEM classroom, see Durham *et al.* (2020).

Instructor Sample. To identify our instructor sample we relied on data from a 2014 census survey distributed to ~1179 alumni of the SI (Aragón *et al.*, 2017). Instructor respondents (*n* = 728) self-reported use of 19 scientific teaching techniques consistent with a taxonomy of scientific teaching practices (Couch *et al.*, 2015) relating to active learning, formative assessment, and inclusive teaching practices. From this 2014 data set, we identified a subgroup of "high-implementing" instructors, defined as those who reported implementing 17 or more scientific teaching practices in their course(s) since attending an SI. Twenty-five instructors were then randomly selected from this group of high implementers and invited to participate in the current study, and 15 instructors ultimately agreed to participate. Their self-reported teaching practices were confirmed with an in-person visit and teaching observation by a member of the research team (AJC) in each targeted instructor's classroom. One instructor was removed from the study due to the inconsistency between the instructor's self-reported teaching practices data and the researcher's observation data.

Fourteen high-implementing instructors from 14 institutions were ultimately included in the study, including 11 women and three men. Each instructor attended at least one SI training.

Seven instructors taught in R1 public universities, three instructors taught in R1 private universities, two instructors taught in R2 public universities, and two instructors taught in other types of universities.

Student Sample. Students who enrolled in the targeted SI-trained instructors' courses were contacted toward the end of the Spring semester via emails and invited to participate in an online survey administered with Qualtrics survey software (Qualtrics, 2015). Out of the 2787 student surveys sent, 2133 responses were received, for a response rate of 76.5%. Thirteen cases were eliminated from the data due to duplicate responses. We also removed four students who did not consent to participate in the research. Outlier analysis using leverage values, Studentized deleted residuals, and Cook's *D* values indicated 14 extreme cases, which were eliminated from the analysis. Therefore, the final analytic sample includes 2102 unique student responses from 14 high-implementing instructors' classes.

Of the participants, 61.9% identified as female, 32.8% identified as male, and 1.2% identified as "other" or chose not to answer. About 36.5% of the participants were between 18 and 19 years, 40.2% of the participants were between 20 and 21 years, 12.5% of the participants were between 22 and 24 years, 6.0% of the participants were 25 years or older, and 0.8% of the participants chose "other" or not to answer. Most of the participants self-identified as: White non-Hispanic (64.7%), followed by South Asian (7.8%), African American (8.9%), East Asian (7.2%), Hispanic/Latino (7.0%), Middle Eastern (3.6%), and Native American (2.0%). Students were allowed to check multiple racial categories, so the sum of the percentages was greater than 100%.

We converted the variable of race/ethnicity using the following rules: 1) if a student self-identified as African American, Hispanic/Latino, or Native American, we assigned the student to the group of underrepresented minority (URM); 2) among the remaining students, if a student self-identified as South Asian, East Asian, or Middle Eastern, we assigned the student to the group of "Other;" 3) if a student self-identified as only White, we assigned the student to the group of White; and 4) all other cases were considered as missing cases on this variable. Most of the participants were assigned to the group of White (57.8%), followed by Other (16.8%), URM (16.6%), and Missing (8.9%). Students completed the online survey outside

course meeting times and were offered nominal course participation credit (<1% of final grade) for completing the survey. Students in these courses who did not wish to participate in this study were given an alternative assignment worth equivalent credit. This project was granted exempt status from each institution's Institutional Review Board Human Subjects Committee, as it examined standard educational practices. IRBs at other participating institutions were also engaged to follow their institutional human subjects research guidelines.

Measures

Growth Mindset. Students' views on learning as an opportunity for growth were assessed using Dweck and colleagues' three-item mindset measure (Dweck *et al.*, 1995). Sample items included: "You have a certain amount of intelligence, and you can't really do much to change it." Students rated each statement on a scale from 1 ("strongly disagree") to 6 ("strongly agree"); these items were each reverse scored so that higher scores indicated a growth orientation. Growth mindset scale reliability in this study was strong (α = 0.93). Factor loadings of all items ranged from 0.84 to 0.96. The three items of Dweck *et al.*'s (1995) measure were used as observed indicators in the SEM.

Trust. Students' levels of trust in their instructors were assessed using a nine-item measure that examined three components of trust: understanding, acceptance, and care. This measure was developed by Cavanagh *et al.* (2018) to extend Clark and Reis's (1988) close personal relationship framework to educational settings. Students rated their agreement with each of the nine items based on their experiences in their instructor's course in the Spring 2016 term, with the scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"). Sample items included: "My instructor gets me" (understanding subscale); "My instructor accepts me for who I am" (acceptance subscale); and "My instructor truly cares about my educational welfare" (care subscale). Two reversed items were eliminated from the analysis due to low factor loadings. This revised measure showed satisfactory reliability (understanding: α = 0.83; acceptance: α = 0.85; care: α = 0.90) and construct validity (root-mean-square error of approximation [RMSEA] = 0.07, comparative fit index [CFI] = 0.99, the Tucker-Lewis index [TLI] = 0.98). Standardized factor loadings are included in the Supplemental Material (see Supplemental Table S1). To test the SEM, we created four parcels. We define a parcel as an averaged combination of highly correlated items. Each parcel included items from one or two components of trust. For an example of how a parcel was created, see Supplemental Table S2. To test the SEM for the trust variable, we formed four parcels of trust items. Each parcel included items from one or two components of the trust measure (care, understanding, acceptance).

Buy-in: Exposure, Persuasion, Identification, and Commitment (EPIC). Student responses to the introduction of EBPs were assessed using the EPIC framework (Cavanagh *et al.*, 2016), a multidimensional measure that gauges four constructs in relation to experiencing EBPs: exposure (E), persuasion (P), identification (I), and commitment (C). Using a binary "yes"/"no" response scale, students were asked to indicate whether they were exposed to a series of 25 EBPs in their current course (exposure). Sample EBPs items included: "I

answered questions in class using a clicker or other polling method"; "I completed in-class activities in groups of two or more"; and "I presented my scientific ideas in writing."

For the practices to which students indicated exposure, they were then prompted to indicate whether they felt each practice was beneficial (persuasion), whether they liked each practice (identification), and whether they would like to engage in the practice in future courses (commitment).

Student buy-in toward EBPs was indicated by the level of commitment of the EBPs along the exposure–persuasion–identification–commitment pathway. Scores for exposure, persuasion, identification, and commitment from the student sample each ranged from 0 to 1. Given the binary nature of the scale, a Kuder-Richardson reliability coefficient (KR20) was computed for the four dimensions of the instrument (exposure, KR20 = 0.87; persuasion, KR20 = 0.92; identification, KR20 = 0.90; and commitment, KR20 = 0.92), which were each found to show acceptable reliability.

Once a student indicated exposure to an EBP, that student was asked three follow-up questions about persuasion, identification, and commitment. To convert each set of binary data for persuasion, identification, and commitment into continuous data, which would allow us to incorporate them into the SEM model, we created four "parcels" for each construct by averaging six or seven items from the 25 EBPs (25 items divided by four parcels equals six to seven items). We did this because it is impractical to include 25 indicators in the SEM model. For example, the first parcel for persuasion was created based on six items that were randomly selected from the list of 25 EBPs (see Supplemental Table S2). The score was computed using a sum of "yes" responses toward persuasion divided by the sum of "yes" responses toward exposure and represents an uptake of practice (i.e., a ratio of degree of uptake). The same rule was applied to computing the scores of the other three parcels. An illustration of how we created the first indicators of persuasion, identification, and commitment is included in the Supplemental Material (see Supplemental Table S2). The measurement model with the continuous indicators demonstrated a satisfactory model fit (RMSEA = 0.09, CFI = 0.95, TLI = 0.93) and strong factor loadings (see Supplemental Table S1).

Engagement. Students self-reported engagement in self-regulated learning behaviors for the current course using subscales from the Motivated Strategies for Learning Questionnaire (MSLQ) (hereafter engagement; Pintrich *et al.*, 1993), for which participants rated a series of 29 learning behaviors. Student responses were based on the extent to which each item reflected their experiences in their current course. Scale points ranged from 1 ("not at all true of me") to 7 ("very true of me"), with higher scores indicating more substantial engagement. Subscales from the MSLQ included in the current study are: Elaboration (e.g., "When reading for this class, I try to relate the material to what I already know"), Organization (e.g., "When I study for this course, I go through the readings and my class notes and try to find the most important ideas"), Critical Thinking (e.g., "I often find myself questioning things I hear or read in this course to decide if I find them convincing"), Metacognitive Self-Regulation (e.g., "Before I study new course material thoroughly, I often skim it to see how it is organized"), and Time and Study Environment Management (e.g., "I usually

study in a place where I can concentrate on my course work"). Reliability was acceptable for all subscales: Elaboration (six items, $\alpha = 0.77$), Organization (four items, $\alpha = 0.61$), Critical Thinking (five items, $\alpha = 0.63$), Metacognitive Self-Regulation (10 items, $\alpha = 0.78$), and Time and Study Environment Management (four items, $\alpha = 0.71$).

While the confirmatory factor analysis (CFA) model-fit indices for the measures of trust, mindset, buy-in, and persistence all had satisfactory model fit, for the engagement variable, the goodness-of-fit indices were inadequate (see Supplemental Table S1). However, in the current study, we are mostly interested in assessing student engagement in general and the relationship between engagement and other variables, rather than any of the five specific subconstructs of the MSLQ. Therefore, we created four composite parcels to assess the latent variable of engagement by using the domain-representative approach (Little *et al.*, 2002). Each parcel included items from all subconstructs; however, the number of items varied across the four parcels due to the odd number of items in the MSLQ. The measurement model with four parcels demonstrated excellent model fit: RMSEA = 0.04, CFI = 1.00, TLI = 1.00.

Final Course Grade. Students' final course grades were used as the primary performance outcome. Grades were provided by each instructor, then normalized to a 4.0-scale score in the following manner: "A+"/"A" = 4.0, "A−" = 3.7, "B+" = 3.4, "B" = 3.0, "B−" = 2.7, "C+" = 2.4, "C" = 2.0, "C−" = 1.7, "D+" = 1.4, "D" = 1.0, "D−" = 0.7, "F"/"W" = 0.0. Because there was only one indicator for this construct, final course grade was included in the SEM as an observed endogenous variable.

Science Persistence. The likelihood of student persistence in a science field, both within and following the college experience, was assessed using an established measure known as the Persistence in the Sciences (PITS) instrument (Hanauer *et al.*, 2016). Leveraging existing psychological assessment instruments, the PITS measure is a six-factor structure that includes: project ownership (emotion and content), self-efficacy, science identity, scientific community values, and networking. For these six factors, the instrument has been shown to have strong internal consistency (Cronbach's alpha = 0.96) (Hanauer *et al.*, 2016). In addition, the PITS instrument also has five outcome measures/items focused on intent to persist in the sciences with a scale from 1 ("Strongly Disagree") to 5 ("Strongly Agree"), with higher scores indicating greater intent to persist in science.

In the present study, we included only the five outcome items. These items and their standardized factor loadings in the parentheses are: "Following this course, I intend to enroll in similar courses" (0.49); "I intend to complete a science related undergraduate degree" (0.74); "In the future, I intend to enroll in a science related graduate program" (0.81); "My future career will involve collecting, analyzing, and reporting scientific data" (0.60); and, "In the future, I would like to be a research scientist" (0.44). We removed the first (0.49) and fifth (0.44) items from analysis due to both empirical and theoretical reasons. The first reason was empirical, due to unsatisfactory factor loadings. The second reason was more theoretical and based upon observing the poor factor loadings. Specifically, we hypothesize that the unsatisfactory loadings are due to a mismatch between the context presently being studied and that in which the PITS survey was created. The PITS was initially developed to measure persistence among students enrolled in course-based undergraduate research experiences (CUREs). In CURE courses, students conduct research projects and, in doing so, gain some professional research skills. In retrospect, the two eliminated items were highly specific to the CURE research context and therefore were not well-suited to the population polled in this study, that is, students enrolled in a wide range of college science courses.

After eliminating these items, the remaining items demonstrated acceptable internal reliability ($\alpha = 0.74$) and construct validity. Therefore, these three items were used as observed indicators in the SEM.

Analyses

The hypothesized framework (see previous Figure 1) was examined with SEM. In our framework, all constructs except final course grade are treated as latent variables, because, strictly speaking, these variables (e.g., trust) cannot be observed or measured directly. When testing a full SEM, a latent variable analysis yields better estimates of the true relationships between the constructs (Hardre and Reeve, 2003; Hsu *et al.*, 2019). In SEM, parceling is a latent variable measurement practice that is commonly used (Little *et al.*, 2002). A parcel is an aggregate-level indicator that consists of the average of two or more items from the latent variable. Because SEM models with parceled data are more parsimonious, using several parcels for each latent variable can also provide improved reliability and improved relationships with other variables (Little *et al.*, 2002; Brown, 2006). In the current study, parcels were created based on the following rules: 1) four parcels for each construct ensures that the measurement models were overidentified; 2) each parcel includes items from all subconstructs when possible; and 3) items are randomly assigned to parcels.

We used the robust maximum likelihood estimator as the estimation method in our analysis. The following commonly used goodness-of-fit indices were examined to evaluate model fit in the SEM: RMSEA, TLI, and CFI. The TLI and CFI values range from 0 to 1, and values above 0.95 are indicative of a good fit. RMSEA value also ranges between 0 and 1, but values closer to 0 are indicative of a better-fitting model. Values below 0.08 indicate a good fit (Hu and Bentler, 1999). A significant factor loading, as determined by a standardized coefficient of 0.30 or above, indicates that the indicator is a good measure of the underlying factor (Hatcher and Stepanski, 1994). Because the observations in the current study are not independent (i.e., students are "nested" within each instructor), we examined the intraclass correlation coefficients (ICCs) for all endogenous variables (see Supplemental Table S3). The observed indicators of engagement, persuasion, identification, and commitment had small ICCs, suggesting that the impact of the hierarchical structure of the data is minimal. Nevertheless, we found large ICCs in persistence and final course grade. To account for the variance explained by between-instructor differences, the variable (i.e., Faculty ID) was recoded into a set of 13 dummy-coded variables and entered into the SEM to control for nesting (Cohen *et al.*, 2003).

After confirming the hypothesized framework, we carried out a series of multigroup SEMs to evaluate the invariance of path coefficients across White, URM, and Other students and between female and male students. First, we examined the baseline model with all parameters left free to vary across

TABLE 1. Descriptive statistics and bivariate correlation coefficients among variables in the hypothesized framework

	Trust	Growth mindset	Persuasion	Identification	Commitment	Engagement	Science persistence	Course grade
Trust	1	0.07**	0.34**	0.21**	0.21**	0.28**	0.10**	0.18**
Growth mindset		1	0.15**	0.07*	0.11**	0.07*	0.04	0.06*
Persuasion			1	0.12**	0.38**	0.26**	0.11**	0.15**
Identification				1	0.54**	0.19**	0.02	0.10**
Commitment					1	0.16**	0.10**	0.16**
Engagement						1	0.22**	0.17**
Science persistence							1	0.11**
Course grade								1
Mean	3.65	4.42	0.65	0.52	0.42	4.57	3.92	2.66
SD	0.79	1.24	0.31	0.28	0.32	0.86	0.95	0.98
Min	1.00	1.00	0.00	0.00	0.00	1.00	1.00	0.00
Max	5.00	6.00	1.00	1.00	1.00	6.75	5.00	4.00
Skewness	−0.64	−0.60	−0.65	−0.13	0.25	−0.22	−1.14	−0.25
Kurtosis	0.90	−0.35	−0.68	−0.84	−1.19	0.69	1.24	−0.75

*$p < 0.05$.
**$p < 0.01$.

groups. Then we constrained the factor loadings of the measurement to be equal in the groups and compared it with the baseline model. No significant differences in the model fit suggests that the measurement has been understood similarly across groups (Little, 1997). In this study, we used chi-square differences to evaluate the equivalence across groups. After demonstrating the measurement invariance across groups, we moved on to test the invariance of the relations between the constructs across the groups of intererst by constraining the path coefficients to be equal across groups. We constrained all structural links to be equal across groups and compared the all-constrained model with the baseline model. If the change in chi-square between the all-constrained model and the baseline model is nonsignificant, then the equivalence of the structural model is confirmed. Statistics were conducted using R v. 4.0.2 and the lavaan package (Rosseel, 2012).

RESULTS

The Buy-in Framework Was Supported and Can be Used to Understand Students' Learning Experiences in Transformed STEM Classes

Before testing the hypothesized framework, we checked descriptive statistics and zero-order bivariate correlations on all variables (see Table 1). Most of the correlation coefficients were in the predicted direction and significant, which suggested that our data are appropriate for conducting SEM (Tabachnick *et al.*, 2007). Indicator-level correlations and descriptive statistics are reported in Supplemental Table S6.

The test of the buy-in framework using SEM indicated that it was a good fit to the data, $\chi^2(570) = 2748.45$, $p < 0.001$, TLI = 0.92, CFI = 0.93, RMSEA [90% confidence interval (CI)] = 0.046 [0.044, 0.047]. While the χ^2 test was significant, this test is sensitive, especially when the sample size is large, and a significant χ^2 statistic is expected in most SEM models (Brown, 2006). All factor loadings in the measurement model were significant at the $p < 0.001$ level, with associated z-values ranging from 16.65 to 63.57. After verifying the psychometric quality of the measurement model, we evaluated the structural relationships specified in the hypothesized model. All of the completed

standardized path coefficients were significant and in the expected directions.

Figure 2 presents the final structural equation model with completely standardized regression coefficients. In accordance with our hypothesis, student trust ($\beta = 0.37$, $p < 0.001$) and growth mindset ($\beta = 0.15$, $p < 0.001$) were positively associated with persuasion. In line with the EPIC framework, student persuasion was positively associated with student identification ($\beta = 0.15$, $p < 0.001$) and commitment ($\beta = 0.35$, $p < 0.001$). Student identification was positively associated with commitment ($\beta = 0.57$, $p < 0.001$). Student commitment ($\beta = 0.21$, $p < 0.001$) was positively associated with student engagement in self-regulated learning behaviors, which in turn was associated with student persistence in science ($\beta = 0.29$, $p < 0.001$) and final course grade ($\beta = 0.22$, $p < 0.001$).

The Buy-in Framework Was Also Supported across Different Racial/Ethnic Groups and between Male and Female Students

To examine the generalizability of the student buy-in framework across various racial/ethnic groups, we first estimated a model in which all parameters in White, URM, and Other groups were simultaneously and freely estimated. The fit of the initial unconstrained model was good, χ^2 (1710) = 4077.62 $p < 0.001$, TLI = 0.91, CFI = 0.92, RMSEA [90% CI] = 0.048 [0.046, 0.050], which suggested that the buy-in framework represented a good fit to the data in White, URM, and Other groups. Because two of the instructors only had one student in the "Other" group, we only included 11 dummy-coded group-level variables in the invariance analysis regarding race/ethnicity. Model-fit indices for each group can be found in the Supplemental Material (see Supplemental Table S4).

After demonstrating the generalizability of the framework across various racial/ethnic groups, we examined the equivalence of the measurement by constraining the factor loadings of the measurement to be equal in the groups and compared the constrained measurement model with the initial unconstrained model. The fit of the constrained measurement model was
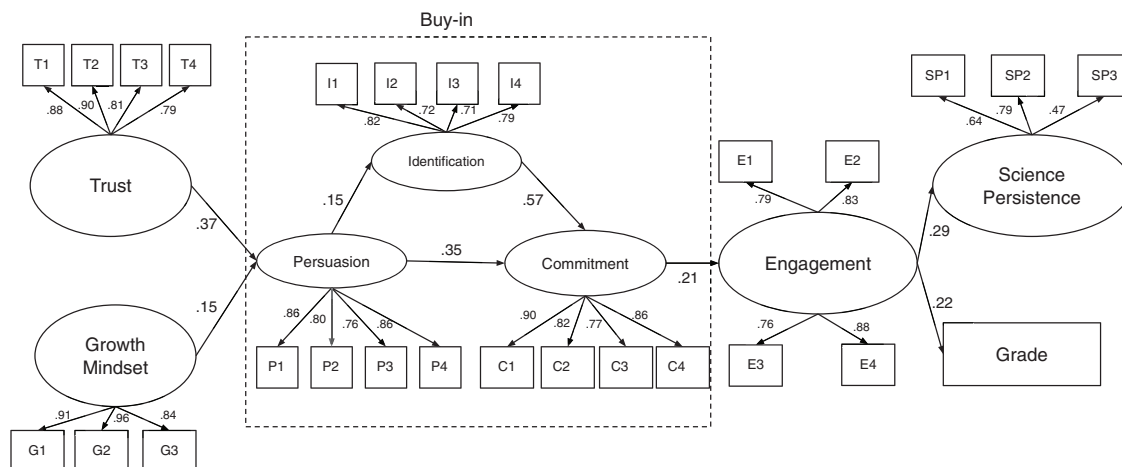
FIGURE 2. The SEMs with standardized coefficients. Model fit: $\chi^2(570) = 2748.45$, $p < 0.001$, TLI = 0.92, CFI = 0.93, RMSEA [90% C. I.] = 0.046 [0.044, 0.047]. Trust item parcels = T1–T4. Growth mindset item parcels = G1–G3. Buy-in item parcels for persuasion = P1–P4. Buy-in item parcels for identification = I1–I4. Buy-in item parcels for commitment = C1–C4. Engagement item parcels = E1–E4. Science persistence item parcels = SP1–SP3.

good, χ^2 (1748) = 4119.44, $p < 0.001$, TLI = 0.91, CFI = 0.92, RMSEA [90% CI] = 0.048 [0.046, 0.050], and the difference between the constrained measurement model and the unconstrained model was nonsignificant, $\Delta\chi^2$ (38) = 41.82, $p = 0.31$, suggesting that all constructs were understood similarly across the three racial/ethnic groups. Next, we constrained all structural links to be equal across the three groups. The fit of the all-constrained model was good, χ^2 (1764) = 4138.55, $p < 0.001$, TLI = 0.91, CFI = 0.92, RMSEA [90% CI] = 0.048 [0.046, 0.049].

The change in chi-square between the all-constrained model and the unconstrained model was nonsignificant, $\Delta\chi^2$ (54) = 60.93, $p = 0.24$, suggesting that all path coefficients were comparable across the three groups. In other words, the relationships specified in the buy-in framework were consistent across ethnic/racial groups. Therefore, the all-constrained model was chosen as the final solution for the invariance analysis regarding ethnicity/race.

The generalizability of the buy-in framework was further examined with respect to gender. The fit of the initial unconstrained model was good, χ^2(1140) = 3336.14, $p < 0.001$, TLI = 0.92, CFI = 0.93, RMSEA [90% CI] = 0.046 [0.044, 0.047], which suggested that the buy-in framework represented a good fit to the data in both male and female groups. Separate model-fit statistics are reported in Supplemental Table S4. Next, we tested the measurement equivalence across the two groups. We constrained the factor loadings of all constructs to be equal across the two groups. The fit of the constrained measurement model was still good, χ^2 (1159) = 3370.86, $p < 0.001$, TLI = 0.92, CFI = 0.93, RMSEA [90% CI] = 0.045 [0.044, 0.047]; however, the difference between the constrained measurement model and the unconstrained model was significant, $\Delta\chi^2$ (19) = 34.72, $p = 0.02$, suggesting that the measures might be understood differently by the two gender groups. With a series of measurement invariance analyses, we found that the factor loadings of measure were different between male students and female students (see Supplemental Table S5). The equivalence of the structural model was not examined with respect to gen-

der due to the nonequivalent measure. We included the correlation matrices and descriptive statistics among subgroups in Supplemental Tables S7 and S8.

DISCUSSION

By testing and validating a student buy-in framework, the present investigation contributes to our understanding of the social and cognitive factors influencing college science students in evidence-based teaching contexts. It directly addresses calls for investigating the underlying mechanisms of how EBPs enhance student outcomes (Dolan, 2015).

We found that students' level of commitment—or buy-in—to EBPs is key to attaining many of the long-desired student outcomes of undergraduate STEM education reform efforts (Woodin *et al.*, 2010), including for students in large introductory science classroom settings. Specifically, we show that students' perceived educational value of their instructors' use of EBPs is positively associated with their level of engagement in course-related activities, their academic performance, and their intent to persist in science. Drawing upon other key elements of the learning process identified in previous studies (Cavanagh *et al.*, 2016; England *et al.*, 2017; Cavanagh *et al.*, 2018; Fink *et al.*, 2018; Tharayil *et al.*, 2018), we examined the factors that might influence students' buy-in to EBPs and their effects on student outcomes. We found that trust and growth mindset influence student buy-in and subsequently affect student academic performance and intent to persist in the sciences. Notably, trust was a particularly strong predictor of student buy-in, over and above that of growth mindset. In addition, our invariance analyses on the buy-in framework showed that: 1) the constructs examined in the study are understood similarly by participants across different gender and racial/ethnic groups, and 2) the relationships specified in the buy-in framework are generalizable to each demographic group studied.

Importantly, the current study moves beyond the context of a single active-learning biology classroom to a number of science courses featuring multiple styles/settings of EBPs. Data from a sample of 2102 undergraduates and 14 instructors teaching

across 28 transformed science classes indicate that the proposed student buy-in framework is valid in understanding college students' learning experiences in general as well as across diverse groups in terms of gender and race/ethnicity. The confirmation of this buy-in framework affirms the need to better understand students' perspectives on course activities as well as the relative importance of social and cognitive processes students go through in order to gauge and meaningfully assess the impact of faculty teaching practices.

For example, although student trust of instructors and growth mindset were both significantly associated with buy-in toward EBPs, we observed that the strength between trust and buy-in was much stronger than that between growth mindset and buy-in. This finding was also reported by Cavanagh and colleagues in a single classroom with one high-implementing EBP instructor (Cavanagh *et al.*, 2018). The importance of this finding here is that across 14 high-implementing instructors, trust of instructors was more than twice a stronger and more consistent predictor of student outcomes—persuasion, identification, and commitment, as well as engagement and course performance—whereas growth mindset had a small effect on student outcomes. Therefore, the pathway from trust to buy-in (i.e., persuasion–identification–commitment) to engagement is a robust association that consistently existed with all 14 of the high-implementing college science instructors and their classrooms. This work reaffirms the significant role that instructors play in enhancing student engagement and outcomes (Fedesco *et al.*, 2019; Snijders *et al.*, 2020).

In the buy-in framework, the variable of trust plays an essential role in fostering student buy-in toward EBPs and raises the question as to what strategies instructors may use to build student trust. Evidence from psychology and education research suggests a number of approaches to promote instructor–student trust relationships within the context of higher education, such as establishing an autonomy-supportive learning environment (Levesque-Bristol *et al.*, 2020), implementing inclusive strategies (Burgstahler, 2015), and having out-of-class communication (Elhay and Hershkovitz, 2019). A next step in line with the current research is to investigate the antecedents of student trust in their instructors within the context of college science EBPs.

LIMITATIONS AND FUTURE RESEARCH

The current study has several limitations we should note. First, this study focused on interpreting undergraduates' learning experiences in STEM courses taught by faculty implementing EBPs with high frequency. While findings provide support for an explanatory buy-in framework of student effects in these environments, questions remain as to the value of this framework in contexts where the implementation of EBPs is more variable. For example, future research should determine whether the buy-in framework holds if there is a more modest implementation of EBPs by a faculty member. Second, the framework tested here examines student buy-in, engagement, course performance, and persistence at a single time point toward the end of the semester across courses. Exploring the interplay among these variables within courses and over time would provide meaningful insight into the potential changes in the strength and direction of these relationships as they evolve over a semester. Third, although the findings of the current study possess the potential for generalizability due to the large and diverse sample of students, our data were mostly collected in biology courses in which 11 of the 14 instructors were female. As such, the sample does not represent all STEM fields and potentially oversamples instructors along gender lines. It would be beneficial to replicate these results with data from other STEM fields. Fourth, although trust of instructors had been found to be significantly associated with students' buy-in toward EBPs, the cross-sectional design used in this study precluded the possibility of demonstrating causal relationships between the variables. Students' buy-in toward EBPs may have contributed to the increase in trust. Accordingly, future researchers may use other research designs, such as a longitudinal research design or a quasi-experimental pre- and posttest design, which could produce more reliable results. Finally, this study did not control for students' prior exposure to EBPs, which may affect their current course experiences. Students' self-reported levels of buy-in could be influenced by prior encounters with the targeted EBPs in other classrooms, influencing their attitudes either positively or negatively.

CONCLUSIONS

This study contributes to research on undergraduate STEM education by providing a framework that depicts how trust and growth mindset increase buy-in to EBPs, which positively influences student outcomes such as engagement in self-regulatory learning strategies, intent to persist in science, and academic performance. This study tested a framework of student buy-in among a multi-institutional sample of undergraduate science students that holds across diverse groups in terms of gender and race/ethnicity. These findings are among the first to detail students' psychological processes, particularly trust, that may impact academic performance and science persistence. Through the lens of the student buy-in framework, we suggest that just training science faculty how to teach in an evidence-based way is an important but insufficient pathway to attain increased student engagement and learning. Instructors should therefore strive to incorporate ways to gain students' buy-in to how the course is being taught and students' trust in their instructors for getting them through a course, especially for students who possess a lower growth mindset or who are from traditionally underrepresented groups.

REFERENCES

American Educational Research Association. (2014). *Standards for educational and psychological testing*. American Educational Research Association, American Psychological Association, National Council on Measurement in Education: Washington, DC.

Aragón, O. R., Dovidio, J. F., & Graham, M. J. (2017). Colorblind and multicultural ideologies are associated with faculty adoption of inclusive teaching practices. *Journal of Diversity in Higher Education*, *10*(3), 201.

Bonem, E. M., Fedesco, H. N., & Zissimopoulos, A. N. (2020). What you do is less important than how you do it: the effects of learning environment on student outcomes. *Learning Environments Research*, *23*(1), 27–44.

Brazeal, K. R., Brown, T. L., & Couch, B. A. (2016). Characterizing student perceptions of and buy-in toward common formative assessment techniques. *CBE—Life Sciences Education*, *15*(4), ar73.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.

Burgstahler, S. (2015). Opening doors or slamming them shut? Online learning practices and students with disabilities. *Social Inclusion*, *3*(6), 69–79.

Cavanagh, A. J., Aragón, O. R., Chen, X., Couch, B. A., Durham, M. F., Bobrownicki, A., … & Graham, M. J. (2016). Student buy-in to active learning in a college science course. *CBE—Life Sciences Education*, *15*(4), ar76.

Cavanagh, A. J., Chen, X., Bathgate, M., Frederick, J., Hanauer, D. I., & Graham, M. J. (2018). Trust, growth mindset, and student commitment to active learning in a college science course. *CBE—Life Sciences Education*, *17*(1), ar10.

Christe, B. L. (2013). The importance of faculty-student connections in stem disciplines. *Journal of STEM Education: Innovations and Research*, *14*(3), 22–26.

Clark, M. S., & Reis, H. T. (1988). Interpersonal processes in close relationships. *Annual Review of Psychology*, *39*(1), 609–672.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Corwin, L. A., Graham, M. J., & Dolan, E. L. (2015). Modeling course-based undergraduate research experiences: An agenda for future research and evaluation. *CBE—Life Sciences Education*, *14*(1), es1.

Couch, B. A., Brown, T. L., Schelpat, T. J., Graham, M. J., & Knight, J. K. (2015). Scientific teaching: Defining a taxonomy of observable practices. *CBE—Life Sciences Education*, *14*(1), ar9.

Degol, J. L., Wang, M.-T., Zhang, Y., & Allerton, J. (2018). Do growth mindsets in math benefit females? Identifying pathways between gender, mindset, and motivation. *Journal of Youth and Adolescence*, *47*(5), 976–990.

Dolan, E. L. (2015). Biology education research 2.0. *CBE—Life Sciences Education*, *14*, ed1.

Durham, M. F., Aragón, O. R., Bathgate, M. E., Bobrownicki, A., Cavanagh, A. J., Chen, X., … & Couch, B. A. (2020). Benefits of a college STEM faculty development initiative: Instructors report increased and sustained implementation of research-based instructional strategies. *Journal of Microbiology & Biology Education*, *21*(2), 21–22.

Dweck, C. S. (2008). *Mindset: The new psychology of success*. New York, NY: Random House Digital.

Dweck, C. S., Chiu, C.-Y., & Hong, Y.-Y. (1995). Implicit theories and their role in judgments and reactions: A word from two perspectives. *Psychological Inquiry*, *6*(4), 267–285.

Elhay, A. A., & Hershkovitz, A. (2019). Teachers' perceptions of out-of-class communication, teacher-student relationship, and classroom environment. *Education and Information Technologies*, *24*(1), 385–406.

England, B. J., Brigati, J. R., & Schussler, E. E. (2017). Student anxiety in introductory biology classrooms: Perceptions about active learning and persistence in the major. *PLoS ONE*, *12*(8), e0182506.

Fedesco, H. N., Bonem, E. M., Wang, C., & Henares, R. (2019). Connections in the classroom: Separating the effects of instructor and peer relatedness in the basic needs satisfaction scale. *Motivation and Emotion*, *43*(5), 758–770.

Fink, A., Cahill, M. J., McDaniel, M. A., Hoffman, A., & Frey, R. F. (2018). Improving general chemistry performance through a growth mindset intervention: Selective effects on underrepresented minorities. *Chemistry Education Research and Practice*, *19*(3), 783–806.

Finkelstein, S., Sharma, U., & Furlonger, B. (2019). The inclusive practices of classroom teachers: A scoping review and thematic analysis. *International Journal of Inclusive Education*, *25*(6), 1–28.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences USA*, *111*(23), 8410–8415.

Graham, M. J., Frederick, J., Byars-Winston, A., Hunter, A.-B., & Handelsman, J. (2013). Increasing persistence of college students in STEM. *Science*, *341*(6153), 1455–1456.

Gross, D., Pietri, E. S., Anderson, G., Moyano-Camihort, K., & Graham, M. J. (2015). Increased preclass preparation underlies student outcome improvement in the flipped classroom. *CBE—Life Sciences Education*, *14*(4), ar36.

Hanauer, D. I., Graham, M. J., & Hatfull, G. F. (2016). A measure of college student persistence in the sciences (PITS). *CBE—Life Sciences Education*, *15*, ar54.

Handelsman, J., Miller, S., & Pfund, C. (2007). *Scientific teaching*. New York, NY: Macmillan.

Hardre, P. L., & Reeve, J. (2003). A motivational model of rural students' intentions to persist in, versus drop out of high school. *Journal of Educational Psychology*, *95*(2), 347–356.

Hatcher, L., & Stepanski, E. J. (1994). *A step-by-step approach to using the SAS system for univariate and multivariate statistics*. Cary, NC: SAS Institute.

Hsu, H. C. K., Wang, C. V., & Levesque-Bristol, C. (2019). Reexamining the impact of self-determination theory on learning outcomes in the online learning environment. *Education and Information Technologies*, *24*(3), 2159–2174.

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

Kim, Y. K., & Sax, L. J. (2014). The effects of student–faculty interaction on academic self-concept: Does academic major matter? *Research in Higher Education*, *55*(8), 780–809.

Komarraju, M., Musulkin, S., & Bhattacharya, G. (2010). Role of student–faculty interactions in developing college students' academic self-concept, motivation, and achievement. *Journal of College Student Development*, *51*(3), 332–342.

Levesque-Bristol, C., Richards, K. A. R., Zissimopoulos, A., Wang, C., & Yu, S. (2020). An evaluation of the integrative model for learning and motivation in the college classroom. *Current Psychology*, 1–13.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *32*(1), 53–76.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*(2), 151–173. doi: 10.1207/S15328007SEM0902_1

Nguyen, K., Husman, J., Borrego, M., Shekhar, P., Prince, M., Demonbrun, M., & Waters, C. (2017). Students' expectations, types of instruction, and instructor strategies predicting student response to active learning. *International Journal of Engineering Education*, *33*(1), 2–18.

Patrick, L. E. (2020). Faculty and student perceptions of active learning. In Mintzes, J. J., & Walter, E. M. (Eds.), *Active learning in college science*. Cham, Switzerland: Springer.

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*(3), 801–813.

Qualtrics. (2015). *Qualtrics survey software*. Provo, UT.

Reis, H. T., & Gable, S. L. (2015). Responsiveness. *Current Opinion in Psychology*, *1*, 67–71.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software*, *48*(2), 1–36.

Snijders, I., Wijnia, L., Rikers, R. M., & Loyens, S. M. (2020). Building bridges in higher education: Student-faculty relationship quality, student engagement, and student loyalty. *International Journal of Educational Research*, *100*, 101538.

Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., ... & Young, A. M. (2018). Anatomy of stem teaching in North American universities. *Science*, *359*(6383), 1468–1470.

Starr, C. R., Hunter, L., Dunkin, R., Honig, S., Palomino, R., & Leaper, C. (2020). Engaging in science practices in classrooms predicts increases in undergraduates' STEM motivation, identity, and achievement: A short-term longitudinal study. *Journal of Research in Science Teaching*, *57*(7), 1093–1118.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). Boston: Pearson.

Tharayil, S., Borrego, M., Prince, M., Nguyen, K. A., Shekhar, P., Finelli, C. J., & Waters, C. (2018). Strategies to mitigate student resistance to active learning. *International Journal of STEM Education*, *5*(1), 7.

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., ... & Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences USA*, *117*(12), 6476–6483.

Woodin, T., Carter, V. C., & Fletcher, L. (2010). *Vision andChangeinBiology Undergraduate Education, A CallforAction*—initial responses. *CBE—Life Sciences Education*, *9*(2), 71–73.

Yeager, D. S., Walton, G. M., Brady, S. T., Akcinar, E. N., Paunesku, D., Keane, L., ... & Dweck, C. S. (2016). Teaching a lay theory before college narrows achievement gaps at scale. *Proceedings of the National Academy of Sciences USA*, *113*(24), E3341–E3348.