Motivational Climate Predicts Student Evaluations of Teaching: Relationships Between Students' Course Perceptions, Ease of Course, and Evaluations of Teaching

Brett D. Jones (iD)
Yasuo Miyazaki (iD)
Mengyun Li
Stephen Biscotte
Virginia Tech

Student evaluations of teaching (SETs) are important at most colleges and universities. One purpose of this study was to determine the extent to which motivational climate was associated with SETs. Another purpose was to determine whether course ease was associated with SETs. Participants included 2,949 undergraduate students from 30 courses at a large public university. Using hierarchical linear modeling, we examined the extent to which students' motivation-related course perceptions (empowerment/autonomy, usefulness, success expectancies, situational interest, and caring) related to SETs at the student and class levels. SETs were highly associated with motivational climate. Furthermore, easier courses were rated lower by students when controlling for motivational climate and the demographic composition of the class. These findings highlight the association between the motivational climate and SETs and suggest that one way to improve SETs may be for instructors to focus on improving the motivational climate rather than making the course easier.

Keywords: *class climate, course evaluation, MUSIC model of motivation, student evaluations of teaching, student motivation, grading leniency*

Student evaluations of teaching (SETs) are important at many colleges and universities because they are often used as an indicator of teaching quality to make important personnel decisions, such as annual reviews, merit raises, and hiring and promotion decisions (Linse, 2017; Miller & Seldin, 2014; Stroebe, 2020; Wachtel, 1998). Another purpose of SETs is to provide instructors with feedback about students' perceptions of their teaching, presumably so they can improve their teaching strategies if needed. Students base their ratings on a variety of factors, including their instructors' teaching strategies (Carpenter et al., 2016) and the motivational climate that the instructor establishes (Griffin, 2016). For example, students' interest in class activities is correlated with their SETs (Jones, 2019). Therefore, despite the fact that SETs are not a good indicator of student learning and can be biased (Carpenter et al., 2020; Peterson et al., 2019; Uttl et al., 2017), they can provide an indication of the motivational climate that is created, in part, by the instructor. By *motivational climate*, we are referring to the aspects of the psychological environment that affect students' motivation and engagement within a course. Researchers have only begun to understand which aspects of the motivational

climate are most related to SETs and the relative importance of these aspects.

One purpose of this study was to determine the extent to which different motivation-related course perceptions were associated with SETs. Although some studies have examined these relationships, they have typically only included two or three motivation-related perceptions (e.g., Filak & Sheldon, 2003; Griffin, 2016), with a few studies including a wider breadth of motivation-related perceptions to assess the motivational climate (e.g., Jones, 2010, 2019; Wilkins et al., 2021). Understanding the extent to which different motivation-related perceptions are associated with students' SETs could be useful to instructors because strategies that have been shown to affect students' motivation-related course perceptions could also be used to increase SETs. For example, if students' perceptions of the usefulness of a course are related to SETs, instructors could implement usefulness strategies to increase both their SETs and improve the motivational climate in the course. And, attending to students' motivation-related perceptions could have positive downstream effects on students' academic motivation and behaviors (Oppenheimer & Hargis, 2020; Serra & McNeely,

2020). For example, a student in a general education course who rates an instructor highly and enjoys the course may be more likely to select courses in that discipline in the future or to seek out more information about that discipline outside of school.

A second purpose of this study was to determine whether the ease of a course is related to SETs when controlling for students' motivation-related class perceptions. Some researchers have found that when instructors provide higher grades, they are more likely to receive higher SETs (Greenwald & Gillmore, 1997), which allows students to shape faculty behavior by encouraging faculty to provide higher grades to receive higher SETs (see Stroebe, 2020, for a discussion). However, the extent to which this finding is causal remains uncertain and alternate explanations are also possible (e.g., Marsh & Roche, 1997; McKeachie, 1997). For example, it may be the case that quality instruction leads to increased learning and achievement, which leads to higher ratings. If "easy" courses are not related to SETs, instructors may be more likely to design rigorous courses without fear of receiving lower SETs, and people making personnel decisions based on SETs may be less likely to attribute high SETs to easy courses rather than other factors, such as teacher effectiveness or motivational climate.

Predictors of Student Evaluations of Teaching

Researchers have correlated SETs with a variety of variables, but likely none so hotly debated as course grades. Students' grades in a course and their SETs tend to be slightly to moderately positively correlated (Brockx et al., 2011; Spooren et al., 2013) and researchers have explained this relationship in a variety of ways. One possibility is that SETs are a valid measure of teaching quality and students rate instructors higher when they earn higher grades because they have learned more due to the instructor's effective teaching (referred to as the validity hypothesis by Marsh, 1984). Another possibility, referred to as the "grading leniency hypothesis" by Marsh (1984), states that when teachers give higher-than-deserved grades, students reward them by giving them higher-than-deserved SETs. Studies have supported both of these possibilities, with some researchers providing evidence that seems to support the validity hypothesis (Arnold, 2009; Centra, 2003; Remedios & Lieberman, 2008) and others providing evidence for the grading leniency hypothesis (Greenwald & Gillmore, 1997; Krautmann & Sander, 1999; McPherson, 2006). After reviewing studies supporting these hypotheses, Brockx et al. (2011) stated, "It is possible to conclude that there is no consensus regarding the interpretation of the relationship between course grades and SET" (p. 292).

Other variables, such as teacher, student, and course characteristics, have also been examined for their relationship with SET scores (for a review, see Brockx et al., 2011).

Some research about teacher characteristics highlights the fact that SETs are based more on factors related to the teacher than to the course (Beran & Violato, 2005; Marsh & Roche, 1997). For example, Marsh (1982) documented that there was virtually no correlation between the SETs of different instructors teaching the same course; yet, the correlation in SETs for the same instructor teaching two different classes was fairly high ($r = .61$ and .72). SETs are also correlated with some important teaching qualities. For example, SETs have been shown to be related to instructors' organization, preparation, enthusiasm, and presentation style (Carpenter et al., 2016; Motz et al., 2017; Serra & Magreehan, 2016; Toftness et al., 2018; Williams & Ceci, 1997), all of which are factors that can affect students' motivation and engagement within courses (Frenzel et al., 2009; Keller et al., 2014; Zhang, 2014). Taken together, these studies indicate that teacher characteristics and teaching strategies are related to SETs.

While these studies have focused on students' ratings of teacher characteristics and qualities, other studies have focused on students' motivation-related course perceptions. These studies have examined how students' motivation-related perceptions of the class environment are related to their SET scores. The term "motivation-related" course perceptions describes student perceptions associated with motivational constructs, such as autonomy (R. M. Ryan & Deci, 2020), utility value (Eccles & Wigfield, 2020), expectancy for success (Bandura, 1997; Eccles & Wigfield, 2020), situational interest (Renninger & Hidi, 2017), and caring/relatedness (R. M. Ryan & Deci, 2020). Students' motivation-related course perceptions are important because they are associated with students' engagement and performance in classes (Christenson et al., 2012; Middleton et al., 2017).

Although researchers have studied motivation-related course perceptions for decades (see Wentzel & Miele, 2016), the extent to which these perceptions are related to SETs is less understood because the research in these two fields has not overlapped significantly, with a few exceptions. For example, SETs have been shown to be related to undergraduate students' perceptions of autonomy support (Demir et al., 2019; Filak & Sheldon, 2003; Griffin, 2016), intrinsic motivation (Filak & Sheldon, 2003; Griffin, 2016), and instructor caring (a.k.a., rapport, relatedness; Demir et al., 2019; Filak & Sheldon, 2003; Perkins et al., 1995). Other researchers have investigated a broader variety of motivation-related perceptions within a study and documented relationships between SETs and autonomy/empowerment, usefulness/utility value, success expectancies, situational interest/intrinsic motivation, and instructor caring (Jones, 2010, 2019; Jones & Skaggs, 2016; Wilkins et al., 2021). In sum, these studies provide evidence that at least five different course perceptions—autonomy/empowerment, usefulness/utility value, success expectancies, intrinsic motivation/interest, and instructor caring—are significantly associated

with SETs. In the next section, we provide more explanation of these course perceptions and discuss their importance.

The Present Study

Although many different class perceptions and motivation constructs exist (see Wentzel & Miele, 2016), we chose to focus on the five course perceptions that comprise the MUSIC Model of Motivation (Jones, 2009, 2018, 2020): perceptions of empowerment/autonomy, usefulness/utility value, success expectancies, situational interest/intrinsic motivation, and caring (the beginning sounds of these five perceptions form the acronym MUSIC). In the MUSIC model, empowerment/autonomy refers to students' perceptions of the amount of control and choice that students have in a course. Usefulness/utility value refers to students' perceptions that the course content is useful to their lives, either currently or in the future. Success expectancies are students' beliefs that they can succeed in the course if they put forth effort. Situational interest is the extent to which students perceive the course to be interesting and enjoyable. Finally, caring refers to students' perceptions that others in the learning environment (i.e., the instructor and other students) care about their learning and well-being.

We included the five MUSIC perceptions in our study for several reasons. First, many studies have shown that the five MUSIC perceptions are distinct (but correlated), including studies with samples of (1) undergraduate students in the United States (Jones et al., 2013; Jones et al., 2014; Jones et al., 2016; Jones & Skaggs, 2016; Wilkins et al., 2021), China (Jones et al., 2017), and Colombia (Jones et al., 2017), and (2) professional school students in the United States (Jones et al., 2019; Pace et al., 2016) and New Zealand (Gladman et al., 2020). Therefore, although these five course perceptions include a range of perceptions, these constructs do not overlap significantly, which reduces redundancy when assessing students' course perceptions. Second, the five MUSIC perceptions can be measured reliably (Jones & Skaggs, 2016; Pace et al., 2016) with instruments that assess constructs common to many current motivation theories, including self-determination theory (R. M. Ryan & Deci, 2020), situated expectancy-value theory (Eccles & Wigfield, 2020), social cognitive theory (Bandura, 1997), interest theories (Renninger & Hidi, 2017), among others (for a more comprehensive list, see Jones, 2018). Third, the five MUSIC perceptions have been shown to be correlated with SETs in prior studies (Jones, 2010, 2019; Jones & Skaggs, 2016; Wilkins et al., 2021) and other researchers have assessed some of these perceptions in their studies with similar results (e.g., Filak & Sheldon, 2003; Griffin, 2016). Fourth, the MUSIC perceptions are malleable in that instructors can adjust their teaching strategies to change students' perceptions. Therefore, our findings could have more direct implications for instructors (e.g., instructors with low SETs should

implement strategies to increase students' perceptions of the usefulness of the content). Fifth, all five MUSIC perceptions have been shown to be important predictors of students' motivation and engagement in courses (Wentzel & Miele, 2016).

We wanted to investigate the relationships between students' MUSIC perceptions and SETs at the student level across a wider variety of courses than had been investigated previously. Some studies have included courses from only one academic discipline, such as education (Griffin, 2016), journalism (Filak & Sheldon, 2003, Study 2), mathematics (Wilkins et al., 2021), and psychology (Demir et al., 2019). Other studies have included a few students from many different courses, which does not allow the results to be generalized to all of the students in those courses (Filak & Sheldon, 2003, Study 1; Jones & Skaggs, 2016). Also, because we were interested in the motivational aspects of the *course* climate, it was important for us to consider class level perceptions in addition to student level perceptions. Examining students' perceptions at the class level allowed us to consider the contextual effects, such as how the dynamics of a teacher and students within a class can affect the overall class climate.

Two research questions guided this study:

Research Question 1: To what extent are students' MUSIC perceptions associated with SETs at the student level and class level?

Research Question 2: To what extent is the ease of the class related to SETs when controlling for students' MUSIC perceptions?

Regarding our first research question, we anticipated that students' MUSIC perceptions would be positively related to SETs at the student and class levels because prior studies have documented these relationships (Filak & Sheldon, 2003; Griffin, 2016; Jones, 2010, 2019; Jones & Skaggs, 2016; Wilkins et al., 2021). Concerning our second research question, we did not have a clear hypothesis because the research findings and interpretations associated with grade leniency are mixed (Greenwald & Gillmore, 1997; Marsh & Roche, 1997; McKeachie, 1997; Stroebe, 2020; Zabaleta, 2007). In our analyses for both research questions, we controlled for three additional variables in our analyses—gender, race/ethnicity, and class size—because researchers have documented that SETs can vary based on these variables (Basow & Martin, 2012; Bavisi et al., 2010; Benton & Pallett, 2013; Ho et al., 2009; Peterson et al., 2019).

Method

Participants, Courses, and Procedure

Participants included 2,949 students from 30 undergraduate courses at a large public university in the Southeastern United

States. Of the 2,949 participants, 1,574 (53.4%) were male, 1,358 (46%) were female, and 17 self-identified as another gender (0.6%). With respect to race/ethnicity, 2,046 (69.4%) of the students were White or Caucasian (not Hispanic), 507 (17.2%) were Asian or Pacific Islander, 131 (4.4%) were more than one race/ethnicity, 113 (3.8%) were Hispanic, 106 (3.6%) were Black or African American, 37 (1.3%) were another race/ethnicity not provided as a survey option, and 9 (0.3%) were Native American.

The 30 courses were part of about 400 undergraduate courses that comprised the general education program, which consisted of the portion of the undergraduate curriculum shared by all students enrolled at the university, regardless of their major. The purpose of the general education program was to provide all students with the opportunity to develop crosscutting skills and capacities and engage with a broad selection of disciplinary fields and perspectives to build a foundation for civic engagement, employability, and lifelong learning (Association of American Colleges and Universities, 2002). Courses that were included in this study met the following criteria: (1) the course instructor participated in an optional, self-selected professional development opportunity designed to help instructors improve their instruction (the purpose of the professional development was to help instructors use assessments effectively and to consider strategies to make their courses more engaging); (2) no more than one section of the course could be included in the study; (3) no instructor could teach more than one course included in the study to ensure that all of the instructors in the study were different; and (4) more than 50% of the students in the course had to have completed the study survey. The courses that met these criteria represented a wide variety of topics, such as theatre, art history, geography, economics, human development, planning and design, computer science, chemistry, and physics (see Table 1 for the complete list). All of the courses were worth three credits except for one course that was worth two credits and one course that was worth four credits.

Students completed an online "Student Perceptions Survey" that they received from their instructor through a URL link provided in an email. The link was sent during the middle half of the course for 21 of the courses and during the last quarter of the course for the other nine courses. The primary purpose of the survey was to obtain feedback from students about the course that instructors could use to improve the course. Typically, the instructors used the feedback to make changes to their course the next semester they taught it. Some instructors included the survey as a graded homework assignment, other instructors gave students time in class to complete it, and others strongly encouraged students to complete it because their feedback could be used to improve the course in the future. To keep students' individual survey responses anonymous to the instructors, a researcher collected all of the survey responses and provided the instructors with the names of students who completed the surveys (if the instructors needed the names to assign grades for the assignment).

This study was approved by the institutional review board at our university, and a student consent form was included as part of the survey. A total of 2,949 students completed the survey and gave their consent to participate in this study. The response rate for each course ranged from 52% to 95%, with a mean value of 72% across courses (see Table 1).

Instruments

Perceptions of the MUSIC Model Components. Students' perceptions of the five MUSIC model components (i.e., empowerment/autonomy, usefulness, success, situational interest, and caring) were measured using the five scales from the MUSIC Model of Academic Motivation Inventory, college student version (MUSIC Inventory; Jones, 2012). The MUSIC Inventory scales measure the extent to which students perceive that they have control of their learning environment in the course (empowerment/autonomy scale; five items), the coursework is useful to their future (usefulness/utility value scale; five items), they can succeed at the coursework (success expectancies scale; four items), the instructional methods and coursework are interesting (situational interest scale; six items), and the instructor cares about whether the student succeeds in the coursework and cares about the student's well-being (caring scale; six items). Students rated all items on a 6-point Likert-format scale with descriptors at each point (1 = *Strongly disagree*, 2 = *Disagree*, 3 = *Somewhat disagree*, 4 = *Somewhat agree*, 5 = *Agree*, 6 = *Strongly agree*). Here is an example item from each scale: "I have the freedom to complete the coursework my own way" (empowerment/autonomy), "In general, the coursework is useful to me" (usefulness), "I am confident that I can succeed in the coursework" (success), "The coursework is interesting to me" (situational interest), and "The instructor cares about how well I do in this course" (caring). Each student received the items in a different, random order. The complete inventory, administration instructions, and validity information are available in the *User Guide* (Jones, 2012). Good Cronbach's alpha values (used as a measure of internal consistency reliability) have been reported in several studies with college students, including Jones and Skaggs (2016; .91 for empowerment, .96 for usefulness, .93 for success, .95 for interest, and .93 for caring); Chittum et al. (2019; .83 for empowerment, .87 for usefulness, .86 for success, .87 for interest, and .82 for caring); and Jones (2019; 37 of the 40 values calculated were between .70 and 1.0). In the present study, the Cronbach alpha values were good to excellent: .86 for empowerment, .94 for usefulness, .87 for success, .92 for interest, and .86 for caring. Some studies have also provided construct and predictive validity evidence for the MUSIC Inventory when it is used

TABLE 1

Participating Courses Ordered by the Number of Students in Each Course

Course topic	No. of students who consented	No. of students in the course	Response rate
Biological engineering	16	17	94%
Systems thinking	16	17	94%
Art and design	17	22	77%
Creative inquiry and design	19	24	79%
Economics	21	27	78%
Geography	24	27	89%
History	27	31	87%
Biology	23	32	72%
Hospitality and tourism management	24	35	69%
Visual arts	26	37	70%
Communication	36	38	95%
History	26	38	68%
Leadership	32	43	74%
Planning and design	50	64	78%
History	47	69	68%
Geography	47	71	66%
Electrical and computer engineering	52	76	68%
Human development	66	79	84%
Environmental science	53	90	59%
Urban affairs and planning	58	108	54%
Animal and poultry science	94	156	60%
Chemistry	185	226	82%
Computer science	136	236	58%
Geography	127	238	53%
Civil and environmental engineering	169	260	65%
Art history	143	277	52%
Geography	342	492	70%
Computer programming	334	530	63%
Theatre	312	579	54%
Physics	427	598	71%
Total	2,949	4,537	

with undergraduate students. For example, factor analysis has been used to provide evidence for the construct validity of the five-factor structure of the MUSIC Inventory in undergraduate courses (Chittum et al., 2019; Jones & Skaggs, 2016; Tendhar et al., 2017). In addition, higher scores on the five MUSIC Inventory scales have been shown to be associated with higher course ratings, instructor ratings, effort (behavioral engagement), and cognitive engagement in undergraduate courses (Chittum et al., 2019; Jones, 2019; Jones & Skaggs, 2016; Wilkins et al., 2021).

Ease of Course. The Ease of Course scale (Jones et al., 2021) assesses the extent to which students perceive the course to be easy. The scale consists of three items that are rated on a 6-point Likert-format scale with descriptors at each point (1 = *Strongly disagree*, 2 = *Disagree*, 3 = *Somewhat disagree*, 4 = *Somewhat agree*, 5 = *Agree*, 6 = *Strongly agree*). The

items are (1) "This course is very easy for me." (2) "I don't need to work my hardest to get a high grade in this course." (3) "In this course, I can get the grade I want with very little effort." This scale has demonstrated acceptable internal consistency reliability in other studies with undergraduate students (α = .73; Jones et al., 2021) and in the present study (α = .82).

Student Evaluations of Teaching. SETs were assessed related to the course and instructor. Students answered one item that measured their overall perceptions of the course and one item that measured their overall perceptions of their instructor. These items were the same as those used in other studies (Jones, 2010; Wilkins et al., 2021) and they are similar to the items on the mandatory course evaluation forms at the participating university. The items were "My overall rating of the *course*" and "My overall rating of the *instructor*

TABLE 2
Univariate Descriptive Statistics for Student Level and Class Level Variables

Level	*M*	*SD*	Min	Max
Student level[a]				
Course rating	4.74	0.96	1.00	6.00
Instructor rating	5.20	0.84	1.00	6.00
Male	0.53	0.50	0	1
Female	0.46	0.50	0	1
Other gender	0.006	0.076	0	1
White	0.69	0.46	0	1
Asian	0.17	0.38	0	1
Other race	0.13	0.34	0	1
Empowerment	4.57	0.87	1.00	6.00
Usefulness	4.42	1.19	1.00	6.00
Success	5.07	0.74	1.00	6.00
Interest	4.49	1.00	1.00	6.00
Caring	5.25	0.63	1.00	6.00
Ease	3.66	1.09	1.00	6.00
Class level[b]				
Class size	98.4	113.42	16	427
Male proportion	0.49	0.20	0.03	0.88
Female proportion	0.51	0.21	0.10	0.97
Other gender proportion	0.01	0.01	0.00	0.05
White proportion	0.73	0.12	0.46	0.92
Asian proportion	0.14	0.10	0.00	0.38
Other race proportion	0.13	0.07	0.00	0.31
Mean course rating	4.72	0.44	3.71	5.44
Mean instructor rating	5.15	0.45	3.90	5.81
Mean empowerment	4.51	0.37	3.81	5.30
Mean usefulness	4.63	0.56	3.43	5.43
Mean success	5.07	0.27	4.48	5.51
Mean interest	4.54	0.55	2.79	5.46
Mean caring	5.32	0.25	4.73	5.75
Mean ease	3.52	0.55	2.44	4.74

[a]$n = 2{,}949$. [b]$n = 30$.

for this course," and both items were rated using the following Likert-format scale: 1 = *Terrible*, 2 = *Poor*, 3 = *Satisfactory*, 4 = *Good*, 5 = *Very good*, and 6 = *Excellent*.

Analysis

We used a two-level multilevel model (hierarchical linear modeling [HLM]; Raudenbush & Bryk, 2002) because the structure of the data was nested, with individual students nested within classes (the Level 1 units were students, and the Level 2 units were classes). HLM allowed us to account for the fact that the responses from the same classes were likely to be more correlated than those in other classes, which creates nonindependent observations that violate the assumption in a multiple linear regression model. The number of macro units in this study (30 classes) met the recommended minimum number of 30 (Hox et al., 2018). The details of our analysis are provided in the appendix, available in the online version of this article.

Results

Descriptive Statistics and Correlations

At the class level, the average number of students in each class (i.e., the cluster size) was 98.4 and ranged from 16 to 427 (see Table 2). Class variables (i.e., Level 2) included the means of the Empowerment (*Mean Empowerment*), Usefulness (*Mean Usefulness*), Success (*Mean Success*), Interest (*Mean Interest*), Caring (*Mean Caring*), and Ease (*Mean Ease*). The control variables at the second level were the proportion of female students and other genders, the proportion of Asian and other races/ethnicities, and the class size.

Bivariate associations among variables at both the student level and class level are reported in Table 3. At the student level, the course rating and instructor rating were highly correlated ($r = .675$). The MUSIC variables were moderately to highly correlated positively with course ratings (r ranged from .419 to .723) and instructor ratings (r ranged from .410 to .623), whereas the ease of course variable had a small, positive correlation with course rating ($r = .125$) and instructor rating ($r = .047$).

At the class level, the mean course rating and instructor rating were very highly correlated ($r = .937$). The mean MUSIC variables were highly positively correlated with course ratings (r ranged from .463 to .902) and instructor ratings (r ranged from .409 to .920). The mean ease of course variable (*Mean Ease*) was not significantly correlated with course rating or instructor rating.

Multilevel Analysis

In this section, we describe the results of the multilevel analysis by summarizing the results of each model tested. We start with the models related to course rating and then move on to the models related to instructor rating.

Course Rating. The results of the analyses for course rating are presented in Table 4. There are six models (Model A to Model F) fitted for the analysis. These are referred to as the taxonomy of statistical models, which is a systematic sequence of models that, as a set, address our research questions (Singer & Willlett, 2003). The ICC for Model A (0.154) is moderately high and indicates that the student responses within the same classes are more similar than those in different classes. In Model B, the independent MUSIC variables

TABLE 3

Pearson Correlation Coefficients for Study Variables at the Student Level and Class Level

Variables	1	2	3	4	5	6	7	8
1. Course rating	—	.675**	.516**	.515**	.469**	.723**	.419**	.125**
2. Instructor rating	.937**	—	.457**	.410**	.416**	.623**	.516**	.047*
3. eMpowerment	.743**	.687**	—	.440**	.491**	.587**	.441**	.239**
4. Usefulness	.463**	.409*	.422*	—	.364**	.664**	.307**	.005
5. Success	.692**	.627**	.735**	.308	—	.493**	.455**	.483**
6. Interest	.902**	.920**	.735**	.505**	.673**	—	.466**	.086**
7. Caring	.691**	.702**	.620**	.423*	.523**	.709**	—	.058**
8. Ease	.163	.097	.442*	−.173	.732**	.152	.017	—

Note. The results for the student level ($n = 2,949$) are shown above the diagonal. The results for the class level ($n = 30$) are shown below the diagonal.
*$p < .05$. **$p < .01$.

are all group mean centered, and therefore, the coefficients represent the unbiased estimates of the student level relationships without considering the ease of the class, student demographics, or class size. All five of the MUSIC variables were statistically significant predictors of course rating and all of the relationships were positive and highly statistically significant ($p \leq .01$; *Empowerment* = 0.055, *Usefulness* = 0.090, *Success* = 0.140, *Interest* = 0.499, and *Caring* = 0.105). In Model C, the student level estimates were the same as those in Model B, which is a virtue of the group mean centering. At the class level (i.e., level 2) for Model C, only *Mean Empowerment* (0.231) and *Mean Interest* (0.579) were significant predictors of course rating. The five MUSIC variables explained 93.0% of the variance in course rating at the class level (see R^2_{L-2} for the pseudo-R^2 row in Table 4).

Model D included only the ease of course variable, which was significantly related to course rating at the student level (0.108), but not the class level (0.127, $p > .05$). Ease of course only explained 2.7% of the variance in course rating at the class level (R^2_{L-2}). Model E combined the MUSIC variables in Model C with the ease of course variable in Model D. Model E was very similar to Model C with respect to the MUSIC variables; however, ease of course was not significantly associated with course rating at the student level and was negatively related to course rating at the class level (−0.349, $p < .001$). Model F added the demographic variables to Model E and all of the MUSIC variables remained statistically significant at the student and class levels, and the ease of course variable remained unrelated to course rating at the student level and negatively related to course rating at the class level (−0.433, $p < .001$). In considering the gender and race/ethnicity variables, the only variable that was significant was that courses with more females tended to be rated higher (0.430, $p < .01$). Larger classes were associated with higher course ratings (0.0004, $p < .05$). Almost all of the variance in course rating (99.6%) at the class level was explained when all of the variables were included in the full model (Model F).

To compare the student level and class level effects more directly, we computed the completely standardized coefficient (β^{**} for level predictor and γ^{**} for Level 2 predictor) and the results are reported in the right-hand column of Table 5 (for examples of half-, semi-, and completely standardized regression coefficients, see Stavig, 1977). These values indicate that the most important variable in predicting course rating is *Success* at the class level (0.699), followed by *Interest* at the class level (0.637), *ease* at the course level (−0.617), *Interest* at the student level (0.519), *Empowerment* at the course level (0.358), *Caring* at the course level (−0.225), *Usefulness* at the course level (−0.181), *Usefulness* at the student level (0.112), *Success* at the student level (0. 110), *Caring* at the student level (0.070), and *Empowerment* at the student level (0.050).

Instructor Rating. The results of the six analyses for instructor rating are presented in Table 6. The ICC for Model A (0.255) is rather high and indicates that the student responses within the same classes are more similar than those in different classes. In Model B, three of the five MUSIC variables were significantly related to instructor rating and all of the relationships were positive and statistically significant ($p < .001$; *Success* = .077, *Interest* = .322, and *Caring* = .383). In Model C, the student level values were the same as Model B and at the class level, only *Interest* (.696, $p < .001$) was a significant predictor of instructor rating. In Model C, the five MUSIC variables explained 88.8% of the variance in instructor rating at the class level.

In Model D, the ease of course variable was not significantly related to instructor rating at the student level or class level. Ease of course only explained 0.6% of the variance in instructor rating at class level. When the MUSIC variables were added to the ease of course variable in Model E, several variables were related to instructor rating at the student level (*Success* = .110, *Interest* = .315, *Caring* = .372, *ease* = −0.044, all $p < .001$) and course level (*Usefulness* = −0.184, $p < .01$; *Success* = .704, $p < .05$; *Interest* = .537, $p < .001$; and *ease* = −0.362, $p < .05$). These

TABLE 4

Results of Fitting a Taxonomy of Multilevel Models for Course Rating

Parameter		Model A	Model B	Model C	Model D	Model E	Model F
Fixed effects							
Predictor	Intercept (γ_{00})	4.719*** (0.075)	4.719*** (0.077)	4.730*** (0.025)	4.718*** (0.074)	4.730*** (0.022)	4.710*** (0.180)
	Empowerment (γ_{10})		0.055** (0.019)	0.055** (0.019)		0.056** (0.019)	0.055** (0.019)
	Usefulness (γ_{20})		0.090*** (0.017)	0.090*** (0.017)		0.089*** (0.017)	0.090*** (0.017)
	Success (γ_{30})		0.140*** (0.020)	0.140*** (0.020)		0.142*** (0.022)	0.143*** (0.023)
	Interest (γ_{40})		0.499*** (0.020)	0.499*** (0.020)		0.498*** (0.020)	0.498*** (0.020)
	Caring (γ_{50})		0.105*** (0.024)	0.105*** (0.024)		0.105*** (0.024)	0.106*** (0.025)
	Ease (γ_{60})				0.108*** (0.017)	−0.003 (0.014)	−0.004 (0.014)
	Female (γ_{70})						−0.007 (0.025)
	Other Gender (γ_{80})						0.027 (0.156)
	Asian (γ_{90})						0.019 (0.033)
	Other Race (γ_{100})						−0.052 (0.035)
	Mean Empowerment (γ_{01})			0.231* (0.110)		0.392*** (0.104)	0.373*** (0.103)
	Mean Usefulness (γ_{02})			−0.032 (0.048)		−0.131* (0.051)	−0.125** (0.040)
	Mean Success (γ_{03})			0.108 (0.136)		0.845** (0.282)	1.000*** (0.212)
	Mean Interest (γ_{04})			0.579*** (0.085)		0.432*** (0.090)	0.447*** (0.080)
	Mean Caring (γ_{05})			0.064† (0.133)		−0.177 (0.137)	−0.348** (0.111)
	Mean Ease (γ_{06})				0.127 (0.137)	−0.349** (0.120)	−0.433*** (0.096)
	Mean Female (γ_{07})						0.430** (0.148)
	Mean Other Gender (γ_{08})						0.387 (2.065)
	Mean Asian (γ_{09})						0.374 (0.254)
	Mean Other Race (γ_{10})						−0.227 (0.342)
	Class Size (γ_{11})						0.0004* (0.0002)
Variance component	Level 1 Within-class ($\sigma^{2\circ}$)	0.819*** (0.021)	0.402*** (0.011)	0.403*** (0.011)	0.808*** (0.021)	0.403*** (0.011)	0.403*** (0.011)
	Level 2 Class mean (τ_{00})	0.149*** (0.044)	0.169*** (0.046)	0.010*** (0.004)	0.145*** (0.042)	0.006* (0.003)	<0.001 (0.001)
	ICC	0.154					
Pseudo-R^2 statistics and goodness of fit	R^2_{L-1}	Base	0.509	0.509	0.013	0.508	0.508
	R^2_{L-2}	Base	−0.134	0.930	0.027	0.960	.996
	Deviance (*N* of estimated parameters)	7852.833 (3)	7777.658 (8)	7717.641 (13)	7811.701 (5)	5709.840 (15)	5690.738 (24)
	AIC	7858.833	7793.658	5743.641	7821.701	5739.84	5738.738
	BIC	7876.801	7841.572	5821.500	7851.647	5829.678	5882.479

Note. p Values are based on *t* tests for fixed effects parameters and *z* tests for variance component parameters; parameters are estimated by full maximum likelihood; Base in pseudo-R^2 indicates that the R^2 was computed based on the value for the model in the column. ICC = intraclass correlation coefficient; AIC = Aikaike information criterion; BIC = Bayesian information criterion.

†$p \leq .10$. *$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

TABLE 5

Half and Completely Standardized Regression Coefficients for Course Rating for Model F

Predictor	Raw coefficient (β)	SDx	SD_Y	Half standardized coefficient, $\beta^* = \beta(SDx)$	Completely standardized coefficient, $\beta^{**} = \beta \frac{SDx}{SD_Y}$
Level 1 predictor (X)					
Empowerment (γ_{10})	0.055**	0.87	0.96	0.048	0.050
Usefulness (γ_{20})	0.090***	1.19	0.96	0.107	0.112
Success (γ_{30})	0.143***	0.74	0.96	0.106	0.110
Interest (γ_{40})	0.498***	1.00	0.96	0.498	0.519
Caring (γ_{50})	0.106***	0.63	0.96	0.067	0.070
Ease (γ_{60})	−0.004	1.09	0.96	−0.004	−0.005

Predictor	Raw coefficient (γ)	$SD(\bar{x})$	$SD(\beta_{0j}) = (\sqrt{\tau_{00}})$	Half standardized coefficient, $\gamma^* = \gamma(SD(\bar{x}))$	Completely standardized coefficient, $\gamma^{**} = \gamma \frac{SD(\bar{x})}{\sqrt{\tau_{00}}}$
Level 2 predictor ($\bar{X}$)					
Mean Empowerment (γ_{01})	0.373***	0.37	0.386	0.138	0.358
Mean Usefulness (γ_{02})	−0.125**	0.56	0.386	−0.070	−0.181
Mean Success (γ_{03})	1.000***	0.27	0.386	0.270	0.699
Mean Interest (γ_{04})	0.447***	0.55	0.386	0.246	0.637
Mean Caring (γ_{05})	−0.348**	0.25	0.386	−0.087	−0.225
Mean Ease (γ_{06})	−0.433***	0.55	0.386	−0.238	−0.617

Note. The expression of $SD(\beta_{0j}) = (\sqrt{\tau_{00}})$ indicates the standard deviation of the Level 2 intercept (β_{0j}) in the unconditional multilevel model (Model A), and it is obtained as the square root of the estimate of the Level 2 variance, τ_{00}, where the estimate of τ_{00} was 0.146.

*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

relationships remained similar at the student level when the demographic variables were added in Model F (see Table 6). In addition, demographic class composition such as identifying as a non-male (*Mean Female* = −0.383, $p < .01$; *Mean Other Gender* = −4.027, $p < .05$), as Asian (*Mean Asian* = −0.859, $p < .001$), or non-White and non-Asian (*Mean Other Race* = −0.714, $p < .05$) was negatively related to instructor rating. Furthermore, instructors of larger classes received higher ratings (0.001, $p < .001$). Almost all of the variance at the class level in instructor rating (99.96%) was explained when all of the variables were included in the full model (Model F).

The results for the completely standardized regression coefficient β are reported in the right-hand column of Table 7. These values indicate that the most important variables in predicting instructor rating were in the following order: *Interest* at the class level (0.890), *Success* at the class level (0.505), *ease* at the class level (−0.484), *Interest* at the student level (0.374), *Caring* at the student level (0.280), *Usefulness* at the class level (−0.122), *Success* at the student level (0.093), and *Ease* at the student level (−0.058).

We examined the possibility of multicollinearity among the predictor variables and determined that multicollinearity did not have a significant effect on the results. The details of our multicollinearity analysis are provided in the online appendix.

Discussion

The primary purpose of this study was to determine the extent to which students' motivation-related class perceptions were associated with their SETs, both at the student level and class level. A secondary purpose was to determine whether the ease of the course was related to SETs when controlling for students' motivation-related class perceptions. In this section, we discuss our findings related to the primary purpose first, followed by those related to the secondary purpose.

Relationships Between Motivation-Related Class Perceptions and SETs

Our first research question was: To what extent are students' MUSIC perceptions associated with SETs at the student level and class level? We measured five motivation-related perceptions (eMpowerment, Usefulness, Success, Interest, and Caring; referred to as "MUSIC perceptions") and assessed SETs using an overall measure of the course and an overall measure of the instructor. We documented that students' MUSIC perceptions were significantly related to their SETs at both the student and class levels, although two of the relationships were negative at the class level.

TABLE 6
Results of Fitting a Taxonomy of Multilevel Models for Instructor Rating

Parameter		Model A	Model B	Model C	Model D	Model E	Model F
Fixed effects							
Predictor	Intercept (γ_{00})	5.155*** (0.076)	5.154*** (0.078)	5.169*** (0.030)	5.154*** (0.076)	5.170*** (0.027)	5.151*** (0.016)
	Empowerment (γ_{10})		0.011 (0.018)	0.011 (0.018)		0.018 (0.018)	0.021 (0.018)
	Usefulness (γ_{20})		0.019 (0.016)	0.019 (0.016)		0.017 (0.016)	0.018 (0.016)
	Success (γ_{30})		0.077*** (0.019)	0.077*** (0.019)		0.110*** (0.021)	0.105*** (0.022)
	Interest (γ_{40})		0.322*** (0.019)	0.322*** (0.019)		0.315*** (0.019)	0.314*** (0.019)
	Caring (γ_{50})		0.383*** (0.023)	0.383*** (0.023)		0.372*** (0.023)	0.373*** (0.023)
	Ease (γ_{60})				0.040 (0.015)	−0.044*** (0.013)	−0.045*** (0.013)
	Female (γ_{70})						−0.029 (0.023)
	Other Gender (γ_{80})						0.157 (0.146)
	Asian (γ_{90})						−0.047 (0.031)
	Other Race (γ_{100})						0.001 (0.033)
	Mean Empowerment (γ_{01})			0.078 (0.134)		0.233 (0.132)	−0.059 (0.090)
	Mean Usefulness (γ_{02})			−0.084 (0.059)		−0.184** (0.065)	−0.087* (0.034)
	Mean Success (γ_{03})			−0.065 (0.164)		0.704* (0.341)	0.748*** (0.180)
	Mean Interest (γ_{04})			0.696*** (0.100)		0.537*** (0.108)	0.647*** (0.071)
	Mean Caring (γ_{05})			0.210 (0.162)		−0.020 (0.171)	0.057 (0.095)
	Mean Ease (γ_{06})				0.070 (0.141)	−0.362* (0.143)	−0.352*** (0.083)
	Mean Female (γ_{07})						−0.383** (0.132)
	Mean Other Gender (γ_{08})						−4.027* (1.841)
	Mean Asian (γ_{09})						−0.859*** (0.221)
	Mean Other Race (γ_{10})						−0.714** (0.309)
	Class Size (γ_{11})						0.001*** (<0.001)
Variance component	Level 1 Within-class ($\sigma^{2\circ}$)	0.611*** (0.016)	0.352*** (0.009)	0.352*** (0.009)	0.610*** (0.016)	0.350*** (0.009)	0.352*** (0.009)
	Level 2 Class mean (τ_{00})	0.160*** (0.045)	0.174*** (0.047)	0.018** (0.007)	0.159*** (0.045)	0.014** (0.005)	<0.001 (<0.001)
	ICC	0.255					
Pseudo-R^2 statistics and goodness of fit	R^2_{L-1}	Base	0.424	0.424	0.002	0.427	0.424
	R^2_{L-2}	Base	−0.088	0.888	0.006	0.913	0.9996
	Deviance (*N* of estimated parameters)	7000.297 (3)	5386.811(8)	5328.250(13)	6992.566 (5)	5310.720(15)	5285.078(24)
	AIC	7006.297	5402.811	5354.25	7002.566	5340.72	5333.078
	BIC	7024.264	5450.724	5432.110	7032.512	5430.558	5476.819

Note. p Values are based on *t* tests for fixed effects parameters and *z* tests for variance component parameters; parameters are estimated by full maximum likelihood; Base in Pseudo-R^2 indicates that the R^2 was computed based on the value for the model in the column.
[†]$p \le .10$. *$p \le .05$. **$p \le .01$. ***$p \le .001$.

At the student level, all five MUSIC perceptions were positively, significantly correlated with the SETs (i.e., course and instructor ratings), with correlations ranging from .410 to .723 (see Table 3). These findings are consistent with the MUSIC Model of Motivation theory (Jones, 2009, 2018, 2020) and research studies that have reported correlations between students' MUSIC perceptions and SETs in undergraduate college courses at the student level (Jones, 2010, 2019; Jones & Skaggs, 2016; Wilkins et al., 2021). Even when controlling for all of the other variables in the model (Model F, Tables 4 and 6), all five MUSIC perceptions were significantly correlated with course rating and three of the MUSIC perceptions (i.e., success, interest, and caring) were significantly correlated with instructor rating. These results are important because they demonstrate that SETs are associated with course perceptions (MUSIC) that have been identified by motivation researchers as being vital to students' engagement in classes and learning activities more generally (Wentzel & Miele, 2016). Although some researchers have investigated these relationships (e.g., Demir et al., 2019; Filak & Sheldon, 2003; Griffin, 2016; Wilkins et al., 2021), the research literature about SETs has remained fairly distinct from the literature about student motivation with little overlap between these literatures. The present study indicates that there are relationships between motivation constructs and SETs that could provide greater insight into the findings in both literatures.

Another contribution of the present study is that we assessed the class level relationships by conducting a multilevel analysis. Class level relationships take into consideration contextual effects, such as the fact that the dynamics of the teacher and students in a class can affect the motivational climate within the class. The models that included the MUSIC perception variables and the control variables (i.e., ease of class, gender, race/ethnicity, and class size; see Model F in Tables 4 and 6) accounted for almost all of the variance at the class level in course rating (99.6%) and instructor rating (99.96%). These findings indicate that the classroom motivational climates (as reflected by the class average perceptions of empowerment, usefulness, success, interest, and caring) predict students' course and instructor ratings. It is important to note that even with only the MUSIC variables in the model (Model C), these variables explained 93.0% of the variance in course rating and 88.8% of the variance in instructor rating at the class level. Similar to the findings at the student level, these class level findings highlight the strong associations between students' motivation-related perceptions and their SETs and add to these findings by demonstrating that the motivational climate of the class is highly related to SETs.

The Importance of Interest and Success. A few patterns emerge when examining the MUSIC perceptions that are most highly predictive of SETs (see Tables 5 and 7). Interest and success at the class level are the highest predictors of course

and instructor ratings. These two *class* level variables are more strongly associated with SETs than any of the other class level or student level variables, which suggests that SETs are strongly associated with the motivational climate of the class.

Because this study is correlational, we cannot conclude that higher class level interest and success perceptions *caused* students to rate their courses and instructors higher. However, given the high correlations between these variables, it would be reasonable to design experimental studies to investigate whether increasing students' perceptions of interest and success leads to higher SETs. In fact, researchers have successfully implemented interventions to increase one or more of students' motivation-related perceptions in a variety of ways (Lazowski & Hulleman, 2016; Lin-Siegler et al., 2016; Reeve et al., 2004). For example, when lower-performing undergraduates were asked to write about how the course material related to their lives, they reported higher usefulness, interest, and expectancy for success in the course (Hulleman et al., 2017). Similarly, undergraduates reported higher levels of usefulness and interest when they were placed in groups of three or four students and asked to discuss: (1) how the topics or assignments in the course related to their goals, (2) which topics or assignments on the syllabus were most interesting to them, and (3) questions they had about the course or instructor (McGinley & Jones, 2014). These types of interventions may be useful in not only increasing students' motivation-related class perceptions but also increasing SETs.

Because the interest variable in the present study measures *situational* interest, it can be triggered by the environment (Renninger & Hidi, 2017). In fact, Hidi and Renninger (2019) note that situational interest is typically triggered by other people or affordances of the environment and that repeated opportunities are needed to sustain interest. The implication is that within a class, teachers can play a critical role in providing experiences that can interest students during class. As a few examples, instructors can provide hands-on activities (Swarat et al., 2012), incorporate games and/or cooperative learning activities (Bergin, 1999), and stimulate emotional arousal (e.g., showing enthusiasm, pacing the lesson appropriately, varying instructional activities; Jones, 2018). Instructors can affect students' perceptions of *success* through a variety of strategies, such as providing explicit expectations (Wang et al., 2018), giving specific and honest feedback (Van den Bergh et al., 2014), designing activities at an appropriate level of difficulty (Shernoff et al., 2003), and helping students understand that their success is related to a combination of their effort and the use of relevant strategies (Dweck, 2006; Weiner, 2000).

Although empowerment perceptions were not as strongly associated with course rating as interest and success perceptions, it may be a variable that is worth manipulating in experimental studies to determine its effect on SETs because it has been shown to be an important predictor of students' engagement in empirical studies (Cheon et al., 2020; Griffin,

TABLE 7
Half and Completely Standardized Regression Coefficients for Instructor Rating for Model F

Predictor	Raw coefficient (β)	SDx	SD_Y	Half standardized coefficient, $\beta^* = \beta(SDx)$	Completely standardized coefficient, $\beta^{**} = \beta \dfrac{SDx}{SDY}$
Level 1 predictor (X)					
Empowerment (γ_{10})	0.021	0.87	0.84	0.018	0.022
Usefulness (γ_{20})	0.018	1.19	0.84	0.021	0.026
Success (γ_{30})	0.105***	0.74	0.84	0.078	0.093
Interest γ_{40})	0.314***	1.00	0.84	0.314	0.374
Caring (γ_{50})	0.373***	0.63	0.84	0.235	0.280
Ease (γ_{60})	−0.045***	1.09	0.84	−0.049	−0.058

	Raw coefficient (γ)	$SD(\bar{x})$	$SD(\beta_{0j}) = (\sqrt{\tau_{00}})$	Half standardized coefficient $\gamma^* = \gamma\left(SD(\bar{x})\right)$	Completely standardized coefficient $\gamma^{**} = \gamma \dfrac{SD(\bar{x})}{\sqrt{\tau_{00}}}$
Level 2 predictor					
Mean Empowerment (γ_{01})	−0.059	0.37	0.40	−0.022	−0.055
Mean Usefulness (γ_{02})	−0.087*	0.56	0.40	−0.049	−0.122
Mean Success (γ_{03})	0.748***	0.27	0.40	0.202	0.505
Mean Interest (γ_{04})	0.647***	0.55	0.40	0.356	0.890
Mean Caring (γ_{05})	0.057	0.25	0.40	0.014	0.036
Mean Ease (γ_{06})	−0.352***	0.55	0.40	−0.194	−0.484

Note. The expression of $SD(\beta_{0j}) = (\sqrt{\tau_{00}})$ indicates the standard deviation of the Level 2 intercept (β_{0j}) in the unconditional multilevel model (Model A), and it is obtained as the square root of the estimate of the Level 2 variance, τ_{00}, where the estimate of τ_{00} was 0.160.

*$p \leq$ .05. **$p \leq$ .01. ***$p \leq$ .001.

2016; Reeve et al., 2020). The fact that empowerment was a predictor of course rating, but not a predictor of instructor rating indicates that students view their empowerment/autonomy as an aspect of the course more than an indication of instructor quality.

The Conundrum of Usefulness and Caring. Although the simple correlations between usefulness and SETs, and caring and SETs, are positive at the class level, their relationships become negative or nonexistent at the class level when these variables are included in the full model with the other variables (usefulness is negatively related to course and instructor rating; caring is negatively related to course rating and unrelated to instructor rating). The reason for this finding was that in the full model, the variance explained in the SETs was shared with all the variables; and thus, only variables that accounted for unique variance in SETs were identified as statistically significant predictors. Because the correlations among students' MUSIC perceptions were statistically significantly correlated, the variance that remains for usefulness and caring after controlling for the other variables is negatively related to the SETs or unrelated (for caring predicting instructor rating). As noted previously, we determined that multicollinearity among the MUSIC perceptions was not a problem statistically; and therefore, we ruled

out the possibility of multicollinearity as a reason for these findings.

We found no theoretical or empirical rationale for why usefulness would be negatively related to SETs when controlled for by the other MUSIC perceptions; consequently, we can only speculate as to why this association exists. Perhaps students reward instructors with higher ratings when courses are perceived as less useful because they understand the difficulties in making seemingly less useful courses relevant to their lives. This can be especially true in the general education courses included in this study because the usefulness may not be as obvious as it is in courses within their major field of study. For example, a history course may not be perceived to be useful by a student majoring in engineering; yet, if the instructor is able to make the history course interesting or make students believe that they can succeed in the course, students may still rate the course highly (as is evidenced by the fact that interest and success were highly related with SETs in this study).

Similarly, we found no research to explain why caring perceptions at the class level were negatively related to course rating in the full model. Clearly, further research is needed to interpret the negative relationships between SETs and usefulness and caring at the class level. To do so, it could be helpful to identify which aspects of the usefulness and

caring constructs overlap significantly with the other MUSIC perceptions and which aspects do not. For example, it may be possible to create subscales within each MUSIC model component and then determine which subscales are more highly correlated with the other MUSIC scales or subscales. As examples, the caring construct has been divided into two subscales (personal and academic) by some researchers (Jones & Wilkins, 2013) and four subscales (promoting interaction, promoting mutual respect, promoting performance goals, and teacher support) by others (A. M. Ryan & Patrick, 2001).

Relationships Between Course Ease and SETs

Our second research question was: To what extent is the ease of the class related to SETs when controlling for students' MUSIC perceptions? To understand the relationships between ease of course and the other study variables, we ran three models: (1) Model D, which included only the ease of course variable and either the course rating (Table 4) or the instructor rating (Table 6); (2) Model E, which added the MUSIC variables to Model D; and (3) Model F, which added the gender, race/ethnicity, and class size variables to Model E (provided in Tables 4 and 6).

Model D showed that students' perceptions of the ease of the course were not related to their course and instructor ratings except that they were positively related to course rating at the student level. This student level finding is consistent with some prior studies (Greenwald & Gillmore, 1997) in which students tended to rate courses higher when the course was perceived as easier. However, another picture emerges when the MUSIC and other variables are included in Models E and F. In Model F (which includes all the variables), ease was negatively correlated with both course and instructor rating at the class level and negatively correlated with instructor rating at the student level (it was uncorrelated with course rating at the student level). These findings indicate that easier courses were rated lower than harder courses once motivation-related variables were controlled for both at the student and class levels, which is a finding that seems to contradict some other studies (Greenwald & Gillmore, 1997; Krautmann & Sander, 1999; McPherson, 2006).

To understand these results, it may be useful to examine the role of success perceptions, which measures students' beliefs that they can succeed at the coursework if they put forth effort. Success perceptions could be high if the course was easy, but they could also be high if the course was hard, such as when students have the resources needed to be successful (e.g., example problems, study guides), the expectations are clear (e.g., clear directions, rubrics are provided), or some other conditions are present that lead students to believe they can succeed even when the course is perceived to be difficult. By including all of the MUSIC variables in the model along with the ease variable, we were able to parse out the individual variance that could be attributed to each variable.

For example, the variance that remains for the ease variable is the part that is unique to ease that does not include the variance that can be attributed to MUSIC perceptions, such as their success perceptions. Given that, SETs were rated lower when courses were perceived as being easy and higher when students had higher MUSIC perceptions. This finding is consistent with others who have reported that higher workloads can lead to higher SETs if students learn more (Marsh & Roche, 2000). Of course, unreasonably high workloads may lead to student frustration and lower SETs (Kulik, 2001), possibly because students perceive the excessive work to be useless or an indication that the instructor does not care about them enough to consider how an excessive workload affects their lives. Future studies could investigate whether it is possible to increase SETs by creating challenging courses and helping students believe that they can succeed by providing appropriate resources and expectations.

The Role of Sex, Race/Ethnicity, and Class Size

We did not have a specific research question related to the role of sex and race/ethnicity primarily because we had no reason to believe that the relationships in this study would vary in any systematic way by gender based on prior studies (Jones, 2010; Nowell, 2007; Zabaleta, 2007). Nonetheless, we used these variables as co-variates in our analysis to determine if they would have any effect on the study variables; and consequently, should be considered as variables to examine in more detail in future studies.

Related to course rating (see Table 4), the only two significant findings were at the course level: (1) classes with a higher percentage of females than males and other genders received higher ratings (0.430, $p \leq .01$) and (2) larger classes were rated higher than smaller ones (0.0004, $p \leq .05$). Related to instructor rating (see Table 6), several findings emerged, all at the course level: (1) instructors of classes with more females and "other" (non-male) genders were rated lower (-0.383, $p \leq .01$ for females' proportion; -4.027, $p \leq .05$ for other genders' proportion), (2) instructors of classes with more non-White students were rated lower (-0.859, $p \leq .001$ for Asian' proportion; -0.714 for Other Race's proportion); and (3) instructors who taught larger classes were rated higher (0.001, $p \leq .001$). Future studies could examine why SETs differ based on these variables.

Limitations and Future Research

Our findings must be interpreted within the context of the study limitations. First, the study included only general education courses and students in these courses may have different motivations and expectations of these courses. In addition, because the instructors of these courses chose to participate in a professional development opportunity to improve their instruction, these instructors may differ from

other instructors in some characteristics (e.g., they may be more committed to improving their teaching). In the future, researchers could compare the results of this study with courses and instructors that were more representative of the population of university courses and instructors. Second, there were some differences in the data collection that were assumed to not have an effect on the results, such as when the surveys were administered during the course and the types of incentives given for completing the survey. Future research could be designed to examine whether these differences in data collection affected the results in any meaningful way. Third, two general "overall" items were used to measure SETs (course rating and instructor rating). Future studies could include more items to assess specific characteristics of teachers and courses. Fourth, there may be other motivation-related class perceptions that are related to SETs that were not included in our study. However, the fact that our full model explained almost all of the class level variance in SETs provides evidence that the variables included were appropriate. Fifth, given that other studies have documented biases in SETs (Carpenter et al., 2020; Peterson et al., 2019), future studies could include instructor-specific class level control covariates (e.g., age, teaching experience) to identify biases against different groups of instructors. Sixth, the fact that fewer than 5% of the students self-identified their race/ethnicity as something other than White/Caucasian or Asian/Pacific Islander impeded our ability to adequately test for the effects of ethnicity/race. Further studies are needed to corroborate our findings in samples of underrepresented students. Seventh, we did not examine the variable associations for nonlinear relationships. Future research could examine whether any of these relationships are nonlinear.

Conclusion

One of our main findings was that the motivational climate of a course—as measured by students' perceptions of empowerment, usefulness, success, interest, and caring within the course—predicts students' overall course rating and instructor rating. This finding suggests that it is reasonable to suspect that changes in one or more MUSIC perceptions could affect students' course and instructor ratings. Because prior studies have shown that instructors have some control over students' MUSIC perceptions (Hulleman et al., 2017; Lin-Siegler et al., 2016; McGinley & Jones, 2014; Reeve et al., 2004), it is possible for future studies to examine how different interventions and teaching strategies can be used to positively affect students' MUSIC perceptions, and consequently, SETs. Experimental studies would be very useful to investigate how intentionally manipulating students' MUSIC perceptions affects their SETs.

Another main finding was that easier courses were rated lower than harder courses when the other study variables

(i.e., MUSIC perceptions, gender, race/ethnicity, and class size) were held constant. For instructors, this finding suggests that attempts to make courses easier in order to improve SETs will likely not be successful. For researchers, these results demonstrate that it is important to control for motivation-related variables (e.g., MUSIC variables) in their studies when they investigate relationships between ease of course and SETs.

Given the results of this study and related studies (e.g., Wilkins et al., 2021), it is becoming clearer that instructors who want to improve their SETs should consider ways to improve the motivational climate of their course. Until more experimental research can be conducted to provide causal evidence of the interactions among motivation-related variables and SETs, the predictive evidence in the current study provides a logical rationale for instructors to focus on increasing students' MUSIC perceptions.

ORCID iDs

Brett D. Jones (iD) https://orcid.org/0000-0003-4002-0084

Yasuo Miyazaki (iD) https://orcid.org/0000-0001-8781-387X

References

Arnold, I. J. M. (2009). Do examinations influence student evaluations? *International Journal of Educational Research*, *48*(4), 215–224. https://doi.org/10.1016/j.ijer.2009.10.001

Association of American Colleges and Universities. (2002). *Greater expectations: A new vision for learning as a nation goes to college*.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.

Basow, S. A., & Martin, J. L. (2012). Bias in student evaluations. In M. E. Kite (Ed.), *Effective evaluation of teaching: A guide for faculty and administrators* (pp. 40–49). Society for the Teaching of Psychology.

Bavisi, A., Madera, J. M., & Hebl, M. R. (2010). The effect of professor ethnicity and gender on student evaluations: Judged before met. *Journal of Diversity in Higher Education*, *3*(4), 245–256. https://doi.org/10.1037/a0020763

Benton, S. L., & Pallett, W. H. (2013, January 28). Class size matters. *Inside Higher Education*. http://www.insidehighered.com/views/2013/01/29/essay-importance-class-size-higher-education

Beran, T., & Violato, C. (2005). Ratings of university teacher instruction: How much do student and course characteristics really matter? *Assessment & Evaluation in Higher Education*, *30*(6), 593–601. https://doi.org/10.1080/02602930500260688

Bergin, D. A. (1999). Influences on classroom interest. *Educational Psychologist*, *34*(2), 87–98. https://doi.org/10.1207/s15326985ep3402_2

Brockx, B., Spooren, P., & Mortelmans, D. (2011). Taking the grading leniency story to the edge: The influence of student, teacher, and course characteristics on student evaluations of teaching in higher education. *Educational Assessment, Evaluation and Accountability*, *23*(4), 289–306. https://doi.org/10.1007/s11092-011-9126-2

Carpenter, S. K., Mickes, L., Rahman, S., & Fernandez, C. (2016). The effect of instructor fluency on students' perceptions of instructors, confidence in learning, and actual learning. *Journal of Experimental Psychology: Applied*, *22*(2), 161–172. https://doi.org/10.1037/xap0000077

Carpenter, S. K., Witherby, A. E., & Tauber, S. K. (2020). On students' (mis)judgments of learning and teaching effectiveness. *Journal of Applied Research in Memory and Cognition*, *9*(2), 137–151. https://doi.org/10.1016/j.jarmac.2019.12.009

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, *44*(5), 495–518. https://www.jstor.org/stable/40197319

Cheon, S. H., Reeve, J., & Vansteenkiste, M. (2020). When teachers learn how to provide classroom structure in an autonomy-supportive way: Benefits to teachers and their students. *Teaching and Teacher Education*, *90*(April), Article 103004. https://doi.org/10.1016/j.tate.2019.103004

Chittum, J. R., Jones, B. D., & Carter, D. M. (2019). A person-centered investigation of patterns in college students' perceptions of motivation in a course. *Learning and Individual Differences*, *69*(January), 94–107. https://doi.org/10.1016/j.lindif.2018.11.007

Christenson, S. L., Reschly, A. L., & Wylie, C. (Eds.). (2012). *Handbook of research on student engagement*. Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-2018-7

Demir, M., Burton, S., & Dunbar, N. (2019). Professor-student rapport and perceived autonomy support as predictors of course and student outcomes. *Teaching of Psychology*, *46*(1), 22–33. https://journals.sagepub.com/doi/10.1177/0098628318816132

Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.

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

Filak, V. F., & Sheldon, K. M. (2003). Student psychological need satisfaction and college teacher-course evaluations. *Educational Psychology*, *23*(3), 235–247. https://doi.org/10.1080/0144341032000060084

Frenzel, A. C., Goetz, T., Lüdtke, O., Pekrun, R., & Sutton, R. E. (2009). Emotional transmission in the classroom: Exploring the relationship between teacher and student enjoyment. *Journal of Educational Psychology*, *101*(3), 705–716. https://doi.org/10.1037/a0014695

Gladman, T., Gallagher, S., & Ali, A. (2020), MUSIC® for medical students: Confirming the reliability and validity of a multi-factorial measure of academic motivation for medical education, *Teaching and Learning in Medicine*, *32*(5), 494–507. https://doi.org/10.1080/10401334.2020.1758704

Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, *52*(11), 1209–1217. https://doi.org/10.1037/0003-066X.52.11.1209

Griffin, B. W. (2016). Perceived autonomy support, intrinsic motivation, and student ratings of instruction. *Studies in Educational Evaluation*, *51*(December), 116–125. https://doi.org/10.1016/j.stueduc.2016.10.007

Hidi, S. E., & Renninger, K. A. (2019). Interest development and its relation to curiosity: Needed neuroscientific research. *Educational Psychology Review*, *31*(4), 833–852. https://doi.org/10.1007/s10648-019-09491-3

Ho, A. K., Thomsen, L., & Sidanius, J. (2009). Perceived academic competence and overall job evaluations: Students' evaluations of African American and European American professors. *Journal of Applied Social Psychology*, *39*(2), 389–406. https://doi.org/10.1111/j.1559-1816.2008.00443.x

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.

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

Jones, B. D. (2009). Motivating students to engage in learning: The MUSIC model of academic motivation. *International Journal of Teaching and Learning in Higher Education*, *21*(2), 272–285. http://www.isetl.org/ijtlhe/

Jones, B. D. (2010). An examination of motivation model components in face-to-face and online instruction. *Electronic Journal of Research in Educational Psychology*, *8*(3), 915–944. https://doi.org/10.25115/ejrep.v8i22.1455

Jones, B. D. (2012). *User guide for assessing the components of the MUSIC® Model of Motivation*. http://www.theMUSICmodel.com

Jones, B. D. (2018). *Motivating students by design: Practical strategies for professors* (2nd ed.). CreateSpace. https://vtechworks.lib.vt.edu/handle/10919/102728

Jones, B. D. (2019). Testing the MUSIC model of motivation theory: Relationships between students' perceptions, engagement, and overall ratings. *Canadian Journal for the Scholarship of Teaching and Learning*, *10*(3), 1–15. https://doi.org/10.5206/cjsotl-rcacea.2019.3.9471

Jones, B. D. (2020). Motivating and engaging students using educational technologies. In M. J. Bishop, E. Boling, J. Elen, & V. Svihla. (Eds.), *Handbook of research in educational communications and technology: Learning design* (5th ed., pp. 9–35). Springer. https://doi.org/10.1007/978-3-030-36119-8_2

Jones, B. D., Byrnes, M. K., & Jones, M. W. (2019). Validation of the MUSIC Model of Academic Motivation Inventory: Evidence for use with veterinary medicine students. *Frontiers in Veterinary Science*, *6*(11), 1–9. https://doi.org/10.3389/fvets.2019.00011

Jones, B. D., Krost, K., & Jones, M. W. (2021). Relationships between students' course perceptions, effort, and achievement in an online course. *Computers and Education Open*, *2*(December), Article 100051. https://doi.org/10.1016/j.caeo.2021.100051

Jones, B. D., Li, M., & Cruz, J. M. (2017). A cross-cultural validation of the MUSIC® Model of Academic Motivation Inventory: Evidence from Chinese- and Spanish-speaking university stu-

dents. *International Journal of Educational Psychology*, *6*(1), 366–385. https://doi.org/10.17583/ijep.2017.2357

Jones, B. D., Osborne, J. W., Paretti, M. C., & Matusovich, H. M. (2014). Relationships among students' perceptions of a first-year engineering design course and their engineering identification, motivational beliefs, course effort, and academic outcomes. *International Journal of Engineering Education*, *30*(6A), 1340–1356. https://www.ijee.ie/contents/c300614A.html

Jones, B. D., & Skaggs, G. E. (2016). Measuring students' motivation: Validity evidence for the MUSIC Model of Academic Motivation Inventory. *International Journal for the Scholarship of Teaching and Learning*, *10*(1), Article 7. http://digitalcommons.georgiasouthern.edu/ij-sotl/vol10/iss1/7

Jones, B. D., Tendhar, C., & Paretti, M. C. (2016). The effects of students' course perceptions on their domain identification, motivational beliefs, and goals. *Journal of Career Development*, *43*(5), 383–397. https://doi.org/10.1177/0894845315603821

Jones, B. D., & Wilkins, J. L. M. (2013). Testing the MUSIC Model of Academic Motivation through confirmatory factor analysis. *Educational Psychology*, *33*(4), 482–503. https://doi.org/10.1080/01443410.2013.785044

Keller, M. M., Goetz, T., Becker, E. S., Morger, V., & Hensley, L. (2014). Feeling and showing: A new conceptualization of dispositional teacher enthusiasm and its relation to students' interest. *Learning and Instruction*, *33*(October), 29–38. https://doi.org/10.1016/j.learninstruc.2014.03.001

Krautmann, A. C., & Sander, W. (1999). Grades and student evaluations of teachers. *Economics of Education Review*, *18*(1), 59–63. https://doi.org/10.1016/S0272-7757(98)00004-1

Kulik, J. A. (2001). Student ratings: Validity, utility and controversy. In M. Theall, P. C. Abrame, & L. A. Mets (Eds.), *New directions for institutional research* (pp. 9–25). Jossey-Bass. https://doi.org/10.1002/ir.1

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education: A meta-analytic review. *Review of Educational Research*, *86*(2), 602–640. http://dx.doi.org/10.3102/0034654315617832

Lin-Siegler, X., Dweck, C. S., & Cohen, G. L. (2016). Instructional interventions that motivate classroom learning. *Journal of Educational Psychology*, *108*(3), 295–299. https://doi.org/10.1037/edu0000124

Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, *54*(September), 94–106. http://dx.doi.org/10.1016/j.stueduc.2016.12.004

Marsh, H. W. (1982). The use of path analysis to estimate teacher and course effects in student ratings of instructional effectiveness. *Applied Psychological Measurement*, *6*(1), 47–60. https://doi.org/10.1177/014662168200600106

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, *76*(5), 707–754. https://doi.org/10.1037/0022-0663.76.5.707

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, *52*(11), 1187–1197. https://doi.org/10.1037/0003-066X.52.11.1187

Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal*

of Educational Psychology, *92*(1), 202–228. https://doi.org/10.1037/0022-0663.92.1.202

McGinley, J., & Jones, B. D. (2014). A brief instructional intervention to increase students' motivation on the first day of class. *Teaching of Psychology*, *41*(2), 158–162. https://doi.org/10.1177/0098628314530350

McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist*, *52*(11), 1218–1225. https://doi.org/10.1037/0003-066X.52.11.1218

McPherson, M. A. (2006). Determinants of how students evaluate teachers. *Journal of Economic Education*, *37*(1), 3–20. https://doi.org/10.3200/JECE.37.1.3-20

Middleton, J., Jansen, A., & Goldin, G. (2017). The complexities of mathematical engagement: Motivation, affect, and social interactions. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 667–699). National Council of Teachers of Mathematics.

Miller, J. E., & Seldin, P. (2014). Changing practices in faculty evaluation: Can better evaluation make a difference? *Academe*, *100*(3), 35–38. https://www.jstor.org/stable/24642931

Motz, B. A., de Leeuw, J. R., Carvalho, P. F., Liang, K. L., & Goldstone, R. L. (2017). A dissociation between engagement and learning: Enthusiastic instructions fail to reliably improve performance on a memory task. *PLoS One*, *12*(7), Article e0181775. https://doi.org/10.1371/journal.pone.0181775

Nowell, C. (2007). The impact of relative grade expectations on student evaluation of teaching. *International Review of Economic Education*, *6*(2), 42–56. https://doi.org/10.1016/S1477-3880(15)30104-3

Oppenheimer, D. M., & Hargis, M. B. (2020). If teaching evaluations don't measure learning, what do they do? *Journal of Applied Research in Memory and Cognition*, *9*(2), 170–174. https://doi.org/10.1016/j.jarmac.2020.03.001

Pace, A. C., Ham, A.-J.L., Poole, T. M., & Wahaib, K. L. (2016). Validation of the MUSIC® Model of Academic Motivation Inventory for use with student pharmacists. *Currents in Pharmacy Teaching & Learning*, *8*, 589–597. https://doi.org/http://dx.doi.org/10.1016/j.cptl.2016.06.001

Perkins, D., Schenk, T. A., Stephan, L., Vrungos, S., & Wynants, S. (1995). Effects of rapport, intellectual excitement, and learning on students' perceived ratings of college instructors. *Psychological Reports*, *76*(2), 627–635. https://doi.org/10.2466/pr0.1995.76.2.627

Peterson, D. A. M., Biederman, L. A., Andersen, D., Ditonto, T. M., & Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. *PLoS One*, *14*(5), Article e0216241. https://doi.org/10.1371/journal.pone.0216241

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage.

Reeve, J., Cheon, S. H., & Jang, H. (2020). How and why students make academic progress: Reconceptualizing the student engagement construct to increase its explanatory power. *Contemporary Educational Psychology*, *62*(July), Article 101899. https://doi.org/10.1016/j.cedpsych.2020.101899

Reeve, J., Jang, H., Carrell, D., Jeon, S., & Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion*, *28*(2), 147–169. https://doi.org/10.1023/B:MOEM.0000032312.95499.6f

Remedios, R., & Lieberman, D. A. (2008). I liked your course because you taught me well: The influence of grades, workload,

expectations and goals on students' evaluations of teaching. *British Educational Research Journal*, *34*(1), 91–115. https://doi.org/10.1080/01411920701492043

Renninger, K. A., & Hidi, S. E. (2017). *The power of interest for motivation and engagement*. Routledge. https://doi.org/10.4324/9781315771045

Ryan, A. M., & Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal*, *38*(2), 437–460. https://doi.org/10.3102/00028312038002437

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*(April), Article 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Serra, M. J., & Magreehan, D. A. (2016). Instructor fluency correlates with students' ratings of their learning and their instructor in an actual course. *Creative Education*, *7*(8), 1154–1165. https://doi.org/10.4236/ce.2016.78120

Serra, M. J., & McNeely, D. A. (2020). The most fluent instructors might choreograph for Beyoncé or secretly be Batman: Commentary on Carpenter, Witherby, and Tauber. *Journal of Applied Research in Memory and Cognition*, *9*(2), 175–180. https://doi.org/10.1016/j.jarmac.2020.02.005

Shernoff, D. J., Csikszentmihalyi, M., Shneider, B., & Shernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly*, *18*(2), 158–176. https://doi.org/10.1521/scpq.18.2.158.21860

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, *83*(4), 598–642. https://doi.org/10.3102/0034654313496870

Stavig, G. R. (1977). The semistandardized regression coefficient. *Multivariate Behavioral Research*, *12*(2), 255–258. https://doi.org/10.1207/s15327906mbr1202_10

Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation: A theoretical and empirical analysis. *Basic and Applied Social Psychology*, *42*(4), 276–294. https://doi.org/10.1080/01973533.2020.1756817

Swarat, S., Ortony, A., & Revelle, W. (2012). Activity matters: Understanding student interest in school science. *Journal of Research in Science Teaching*, *49*(4), 515–537. https://doi.org/10.1002/tea.21010

Tendhar, C., Paretti, M. C., & Jones, B. D. (2017). The effects of gender, engineering identification, and engineering program expectancy on engineering career intentions: Applying hierarchical linear modeling (HLM) in engineering education research. *American Journal of Engineering Education*, *8*(2), 157–170. https://doi.org/10.19030/ajee.v8i2.10072

Toftness, A. R., Carpenter, S. K., Geller, J., Lauber, S., Johnson, M., & Armstrong, P. I. (2018). Instructor fluency leads to higher confidence in learning, but not better learning. *Metacognition and Learning*, *13*(1), 1–14. https://doi.org/10.1007/s11409-017-9175-0

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, *54*(September), 22–42. https://doi.org/10.1016/j.stueduc.2016.08.007

Van den Bergh, L., Ros, A., & Beijaard, D. (2014). Improving teacher feedback during active learning: Effects of a professional development program. *American Educational Research Journal*, *51*(4), 772–809. https://doi.org/10.3102/0002831214531322

Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, *23*(2), 191–212. https://doi.org/10.1080/0260293980230207

Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation*, *24*(3–5), 124–179. https://doi.org/10.1080/13803611.2018.1548798

Weiner, B. (2000). Intrapersonal and interpersonal theories of motivation from an attributional perspective. *Educational Psychology Review*, *12*(1), 1–14. https://doi.org/10.1023/A:1009017532121

Wentzel, K. R., & Miele, D. B. (Eds.). (2016). *Handbook of motivation at school* (2nd ed.). Routledge. https://doi.org/10.4324/9781315773384

Wilkins, J. L. M., Jones, B. D., & Rakes, L. (2021). Students' class perceptions and ratings of instruction: Variability across undergraduate mathematics courses. *Frontiers in Psychology*, *12*, Article 576282. https://doi.org/10.3389/fpsyg.2021.576282

Williams, W. M., & Ceci, S. J. (1997). "How'm I doing?" Problems with student ratings of instructors and courses. *Change: The Magazine of Higher Learning*, *29*(5), 12–23. https://doi.org/10.1080/00091389709602331

Zabaleta, F. (2007). The use and misuse of student evaluation of teaching. *Teaching in Higher Education*, *12*(1), 55–76. https://doi.org/10.1080/13562510601102131

Zhang, Q. (2014). Assessing the effects of instructor enthusiasm on classroom engagement, learning goal orientation, and academic self-efficacy. *Communication Teacher*, *28*(1), 44–56. https://doi.org/10.1080/17404622.2013.839047

Authors

BRETT D. JONES is a professor and leader of the Educational Psychology Program in the School of Education at Virginia Tech. He teaches courses related to motivation, cognition, and teaching strategies, and studies instructional methods that support students' motivation and learning.

YASUO MIYAZAKI is a professor in the School of Education at Virginia Tech. His research involves developing new statistical and measurement modeling methodologies and improving current methodologies for the most pressing issues that plague behavioral and social scientists.

MENGYUN LI is a doctoral candidate in the School of Education at Virginia Tech. She is interested in educational research and evaluation, including instrument design, data collection, analysis, and visualization.

STEPHEN BISCOTTE is the director of general education at Virginia Tech. His scholarship has focused on the non-STEM student experience in STEM general education courses and he gives presentations about innovative structures, university reform efforts, and program evaluation.