Studies in Educational Evaluation 51 (2016) 116–125

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/stueduc

Perceived autonomy support, intrinsic motivation, and student ratings of instruction

Bryan W. Griffin

Department of Curriculum, Foundations, and Reading, Georgia Southern University, P. O. Box 8144, Statesboro, GA 30460, United States

ARTICLE INFO

Article history:
Received 20 January 2016
Accepted 17 October 2016
Available online 28 October 2016

Keywords:
Teacher evaluation
Student evaluation
Student ratings of instruction
Autonomy support
Intrinsic motivation
Instructional evaluation

ABSTRACT

Motivation theory suggests that autonomy supportiveness in instruction often leads to many positive outcomes in the classroom, such as higher levels of intrinsic motivation and engagement. The purpose of this study was to determine whether perceived autonomy support and course-related intrinsic motivation in college classrooms positively predict student ratings of instruction. Data were collected from 47 undergraduate education courses and 914 students. Consistent with expectations, the results indicated that both intrinsic motivation and autonomy support were positively associated with multiple dimensions of student ratings of instruction. Results also showed that intrinsic motivation moderated the association between autonomy support and instructional ratings—the higher intrinsic motivation, the less predictive autonomy support, and the lower intrinsic motivation, the more predictive autonomy support. These results suggest that incorporating classroom activities that engender autonomy support may lead to improved student perceptions of classroom instruction and may also enhance both student motivation and learning.

1. Introduction

For faculty in colleges and universities throughout the world, student ratings of instruction are a common means for evaluating instruction for improvement, and for merit, tenure, and promotion decisions (Chen & Watkins, 2010; Darwin, 2010; Hendry & Dean, 2002; Husbands & Fosh, 1993; Husbands, 1998; Leckey & Neill, 2001; Saroyan & Amundsen, 2001). When instructors consider how they may improve their teaching, the dimensions of instruction found on many instructional rating forms offer some guidance on those areas that one should examine first. For example, it is not uncommon for instructional rating instruments to contain dimensions such as content organization, clarity of presentation, and availability of instructor to students (Abrami, d'Apollonia, & Rosenfield, 2007; Apodaca & Grad, 2005). Feldman (1997) conducted a meta-analysis of research on student ratings of instruction and was able to calculate the mean correlation between various instructional rating dimensions and student achievement. The top four dimensions, all of which demonstrated correlations with achievement ranging from 0.46 to 0.57, were instructor's preparation/organization of course, clarity and understandability of course content, adherence to course objectives, and the perceived outcome or impact of instruction (i.e., skills or knowledge gained).

In addition to the four dimensions listed above, Feldman (1997) also identified two other dimensions that he judged to be of high importance for instruction and learning: instructor stimulates interest in course/subject matter, and instructor motivates students to do their best (or sets high standards for performance). Both of these dimensions correlate 0.38 with student achievement, and both represent important aspects of student motivation in the classroom. Educational researchers have long recognized that motivation plays an important role in student learning (Covington, 2000), and motivation may affect the way students perceive instruction (Feldman, 1998). Within the literature on student ratings of instruction, there is evidence that more motivated students, such as those with higher levels of interest in the subject matter of the course, provide higher ratings when evaluating instructors (Marsh, 1987). Howard and Maxwell (1980), Marsh (1980, 1983), Cashin and Downey (1992), and Prave and Baril (1993) found that students' pre-course interest – defined as desire to take a course – predicted student ratings and course satisfaction. Olivares' (2001) research showed that change in interest over the period of a course, rather than initial pre-course interest, provided an even stronger prediction of instructional ratings.

While pre-course interest predicts student ratings and course satisfaction, this measure does not capture levels of intrinsic or

E-mail address: bwgriffin@GeorgiaSouthern.edu (B.W. Griffin).

http://dx.doi.org/10.1016/j.stueduc.2016.10.007
0191-491X/

extrinsic motivation that may develop because of exposure to either the course material or instructor. Ryan and Deci (2000) explained that intrinsic motivation refers to engagement in an activity because one finds that activity naturally pleasing or interesting, and extrinsic motivation is activity engagement not for internally derived interest but instead for targeted outcomes, rewards, or in reaction to the control of others (Stipek, 1998). Often in student ratings research extrinsic motivation is measured by actual or anticipated grades in a course. The relationship between course grades and student ratings of instruction has been a highly researched topic in this area for decades (Brockx, Spooren, & Mortelmans, 2011; Greenwald & Gillmore, 1997; Marsh & Roche, 1997). Marsh (1987) and Marsh and Roche (2000) noted that, on average, there is a positive association between expected grades and student ratings. As discussed by Greenwald and Gillmore (1997) and Marsh and Roche (1997), what remains unclear is the causal mechanism underlying this relationship. Does the association reflect validity (i.e., better instruction leads to greater understanding and achievement which leads to higher ratings), invalidity (i.e., grading leniency; students rate lower those instructors who are not lenient graders), or a spurious association (i.e., the association between grades and ratings is due to confounding variables, such as motivation)?

Course grades can certainly be a strong motivator for students, but Ryan and Deci (2000) suggested that while various manifestations of extrinsic motivation can be helpful in encouraging learning, one should desire to enhance students' intrinsic motivation as a driving force because of its centrality for self-determined behavior. Teachers can adopt classroom practices that lead to greater student engagement and motivation, noted Niemiec and Ryan (2009), and research shows that classroom instructional activities do predict variations in dimensions of motivation among students (Church, Elliot, & Gable, 2001; Garcia & Pintrich, 1996; Greimel-Fuhrmann & Geyer, 2003; Vansteenkiste et al., 2004). While there are myriad methods teachers may employ to enhance student motivation (Linnenbrink & Pintrich, 2002), the adoption of autonomy supportiveness appears to be one well supported by research and theory (Deci, Vallerand, Pelletier, & Ryan, 1991; Reeve, 2009; Haerens, Aelterman, Vansteenkiste, Soenens, & Van Petegem, 2015).

According to Stefanou, Perencevich, DiCintio, and Turner (2004), autonomy-supportive behavior, as opposed to controlling behavior, incorporates teacher attitudes and actions that are designed to encourage student engagement and learning by including students in decisions, offering choices among academic activities, seeking student input, and providing information and explanations for classroom activities and requirements. By contrast, a controlling instructional style, explained Reeve (2009), is one that may force or pressure students to adopt the teacher's perspective, to think or behave in specific ways, or to rely on extrinsic motivational sources. As noted above, research shows that when teachers adopt an autonomy-supportive instructional approach, a number of positive benefits result including, for example, improved intrinsic motivation and intrinsic goal orientation (Bieg, Backes, & Mittag, 2011; Deci, Nezlek, & Sheinman, 1981; Deci, Schwartz, Sheinman, & Ryan, 1981; Garcia & Pintrich, 1996); enhanced self-regulation and perceptions of efficacy (Williams & Deci, 1996), and higher levels of student engagement (Assor, Kaplan, & Roth, 2002; Haerens et al., 2015; Reeve, Jang, Carrell, Jeon, & Barch, 2004). Reeve (2009) offers a more complete list of empirically supported benefits of autonomy-supportive classroom behaviors.

Given that autonomy-supportive instructional approaches appear associated with more engagement and motivation among students, it is possible that students would perceive autonomy-supportive teachers as better instructors when compared to their more controlling colleagues. Authors of at least three studies have examined the possible link between autonomy-supportive classroom behavior and student ratings of instruction. Filak and Sheldon (2003) reported results of two studies. In their first, Filak and Sheldon asked students to recall a recent course that was important to them and complete a data collection instrument based upon their recollections. Filak and Sheldon found that perceptions of autonomy and competence (i.e., enjoyment of challenge, accomplishment, and stimulation) were the best predictors of both instructor and course ratings. Their second study employed student data from 12 sections of a common course. As before, both autonomy and competence were the strongest predictors of student ratings. In a related study, Filak and Sheldon (2008) found that autonomy support again predicted both course and instructor ratings thus replicating their earlier findings. In addition, they examined a latent variable model and found autonomy support predicted student self-determined motivation and student need satisfaction, and student need satisfaction, in turn, predicted both course and instructor ratings. It is unclear whether Filak and Sheldon (2008) tested a direct path between autonomy support and course or instructor ratings.

More recently, Jones (2010) examined how components of an academic motivation model predict student effort, instructor ratings, and course ratings. Jones collected data from two sections of one course; one section was on-line and the other was face-to-face. Jones' analyses consisted of 12 stepwise regression models, one for each sub-group as defined by student sex (female vs. male) and course section (on-line vs. face-to-face), and for each of three outcomes: student effort, instructor ratings, and course ratings. Jones found mixed results for each outcome modeled; however, the single best predictor across all combinations was situational interest (a temporary and context specific conceptualization of interest), which was a statistically significant predictor in 10 of the 12 models examined. Results showed that the greater situational interest, the greater student effort, the higher instructor ratings, and the higher course ratings. Academic caring (degree to which instructor provides academic support) and autonomy support (which Jones labeled empowerment) were the next most consistent predictors with both showing an association with the modeled outcomes in 5 of the 12 regression models examined. Jones argued that given these results instructors should find ways to generate situational interest among students to enhance effort and possibly alter instructor and course ratings. While not examined in Jones' data, it is possible that autonomy support may have an indirect linkage – mediated through situational interest – to student effort, instructor ratings, and course ratings.

In summary, the research cited above links autonomy support and intrinsic motivation, and both appear to be associated with better instruction as judged by students. Intrinsically motivated students are likely to have a more positive experience within the classroom and therefore should rate higher both instructor and course. Similarly, students who experience more autonomy support should similarly rate higher their instructor and course. The purpose of this study was to examine how intrinsic motivation and perceived autonomy support predict multiple dimensions of student ratings of instruction while simultaneously controlling for a number of covariates, described below, previously demonstrated to predict student ratings.

2. Method

2.1. Participants

Participants in this study included 914 undergraduate students enrolled in 47 randomly selected, education-related courses at a medium-sized (20,000 students), regional university in the

southeastern United States. The data were collected over a period of five semesters. With this sampling strategy it is possible some students were sampled more than once from two or more courses, although the frequency of this occurring is likely to be very low.

The classes ranged in size from 6 to 34 students. Undergraduate education students at this institution are predominately White (~70%) and female (~80%). Most respondents (~74%) reported grade point averages in the range of 2.5–3.5 on a 4.0 scale. Most education-related courses offered at this institution are at the junior and senior level with only a few select courses offered at the sophomore level, therefore most participants in this sample were juniors or seniors.

2.2. Measurement

Since the instrument employed in this study included a large number of items to measure various constructs, it was important that the number of items included be limited to the extent possible so little class time would be consumed during administration. To keep the instrument short, the number of student rating items included was limited to 10. Research on student ratings of instruction shows teaching to be an activity that varies across multiple dimensions (Abrami et al., 2007; Marsh, 1987; Murray, 1997), with some dimensions correlating more strongly with student achievement than others (Feldman, 1997). Drawing from the sources cited above, items were developed or selected to measure eight dimensions of instruction, and two items were used to gather an overall assessment of both instructor and course. For example, the first two items listed below were developed and included because Feldman (1997) shows that each correlates highly with student achievement. The items were preparation and organization of course materials ($r = 0.57$), and clarity in which course material is presented ($r = 0.56$). A number of items were taken from Marsh's (1987) Students' Evaluation of Educational Quality (SEEQ) instrument, or from wording provided by Feldman (1997). Wording for the 10 student rating items is presented in Table 1. Students responded to each of the items using one of two scaling options. For the first eight items the scale ranged from 1 "strongly disagree" to 5 "strongly agree" and for the remaining two items, overall course and overall instructor statements, the scaled ranged from 1 "Poor" to 5 "Excellent."

It is important to demonstrate, for validity purposes, that scores produced by these ten items behave similarly to ratings produced by other valid measures of student evaluations of instruction. Provided below are several examples to demonstrate that these ten items produced scores that behave in ways consistent with prior research. For example, Marsh and Roche (2000) reported, after examining several studies, that the average correlation between overall ratings of instruction and expected course grades to be about 0.20; Feldman (1997) reports that the typical correlation between ratings and expected grades is usually between 0.10 and 0.30. Consistent with these findings, the correlation between ratings and grades for this study ranged from 0.08 to 0.22 with 7 of the 10 correlations between 0.16 and 0.22. Additionally, researchers have examined the relation between class size and student ratings (e.g., Benton & Cashin, 2012; Marsh, 1987) and results generally show no relationship or weak relationships that are typically negative. The correlations between class size and ratings in this study ranged between −0.09 and 0.06. Abrami et al. (2007) reported a factor analysis of correlations among items in Marsh's (1987) SEEQ instrument created a clear one-factor solution that explained about 60% of the variance. Similarly, a factor analysis of scores from these 10 items show one clear factor (first eigenvalue = 6.56, next largest eigenvalue = 0.91) that explains 65% of the item variance with factor loading ranging from 0.70 to 0.87. Hoyt and Lee (2002) provide technical information for the Individual Development and Educational Assessment (IDEA) student ratings of instruction scale used by many universities in the USA. The first 20 items of this scale assess methods of teaching, and correlations among these items range from 0.19 to 0.90 with 84% of these correlations between 0.39 and 0.78. The mean correlation among the IDEA items is 0.65. The ten items used in this study produced correlations that ranged from 0.39 to 0.78 with a mean correlation of 0.63. Taken together, the statistical results reported above suggest the items employed in this study produce scores that behave consistently with ratings reported by other researchers.

Perceived autonomy support was measured by student responses to three statements, "The instructor was willing to negotiate course requirements with students," "Students had some choice in course requirements or activities that would affect their grade," and "The instructor made changes to course requirements or activities as a result of student comments or concerns." These

Table 1
Summary of Outcome and Predictor Variables.

Outcome Measures: Instructional Rating Items	Variable Scale Range
1. The instructor presented the material in a clear and understandable manner.	1 = Strongly Disagree, 5 = Strongly Agree
2. Course materials were well prepared and organized.	1 = Strongly Disagree, 5 = Strongly Agree
3. Students were invited to share their ideas and knowledge.	1 = Strongly Disagree, 5 = Strongly Agree
4. The instructor made students feel welcome in seeking help/advice in or outside of class.	1 = Strongly Disagree, 5 = Strongly Agree
5. Methods of evaluating student work were fair and appropriate.	1 = Strongly Disagree, 5 = Strongly Agree
6. The instructor seems to have a real interest in and concern for students.	1 = Strongly Disagree, 5 = Strongly Agree
7. The instructor gave students useful/helpful feedback on work.	1 = Strongly Disagree, 5 = Strongly Agree
8. The instructor is very knowledgeable in the subject of this course.	1 = Strongly Disagree, 5 = Strongly Agree
9. Overall, how would you rate this course?	1 = Poor, 5 = Excellent
10. Overall, how would you rate this instructor?	1 = Poor, 5 = Excellent
Predictor Variables	
11. Perceived Autonomy Support ($\alpha = 0.85$)	1 = Strongly Disagree, 5 = Strongly Agree
12. Intrinsic Motivation ($\alpha = 0.80$)	1 = Strongly Disagree, 5 = Strongly Agree
13. Extrinsic Motivation with focus on course grades	1 = Strongly Disagree, 5 = Strongly Agree
14. Instructor Grading Leniency	1 = Strongly Disagree, 5 = Strongly Agree
15. Course Difficulty	1 = One of Easiest, 5 = One of Most Difficult
16. Course Workload	1 = Very light, 5 = Very heavy
17. Negative Instructor Reputation	1 = Negative, 0 = Other
18. Positive Instructor Reputation	1 = Positive, 0 = Other
19. Expected Grade	1 = F, 13 = A+
20. Negative Grade Discrepancy	1 = Grade Discrepancy, 0 = other
21. Instructor Sex	1 = male, 0 = female
22. Class Size	6–34 students

items were based partially on an autonomy measure presented in Garcia and Pintrich (1996). The response scale ranged from 1 ("strongly disagree") to 5 ("Strongly agree"). Using data from this current study, internal consistency for these three items was $\alpha = 0.85$. Intrinsic motivation in the course was assessed by responses to the following three statements: "You are very interested in the subject matter covered in this course," "The most satisfying thing for you in this course is trying to understand the content as thoroughly as possible," and "In a class like this, you prefer course material that really challenges you so you can learn new things." These items were adapted from Pintrich, Smith, Garcia, and McKeachie's (1991) measures of task value and intrinsic goal orientation. Responses to these items produced a Cronbach's alpha of 0.80. Lastly, to measure extrinsic motivation with a focus on course grades, the following item was included: "If you had to choose between having a very good understanding of the material in this course, or having a very good grade, getting a very good grade would be your choice." Responses followed the same scale as described above for autonomy support and intrinsic motivation. Results of an exploratory factor analysis of responses to the six items measuring autonomy support and intrinsic motivation showed two distinct factors with eigenvalues of 2.71 and 1.75 (the next largest was 0.49) which together explained 74% of the item variance, and each item loaded distinctly on the factor for which it was designed. In addition, the correlation between intrinsic motivation scores and responses to the extrinsic, grade-oriented motivation item was -0.24 ($p < 0.001$) which provides evidence that the two measures appear to be capturing different motivational constructs. In a study of 3rd through 8th grade students, Lepper, Corpus, and Iyengar (2005) also reported a correlation of -0.24 between measures of intrinsic and extrinsic motivation; the similarity of correlations found in their study and in this study, despite using different measures and different populations, offers additional evidence of construct validity.

In addition to perceived autonomy and intrinsic motivation, students responded to a number of items designed to measure variables related to student ratings. These variables served as covariates in the analyses presented below. Olivares (2001) and Griffin (2004) found that grading leniency predicted ratings. To assess grading leniency, students responded to the following item: "This instructor is a lenient/easy grader" (responses ranged from 1 "strongly disagree" to 5 "strongly agree"). There is debate among researchers about the possible link between student ratings and both course difficulty and course workload (Greenwald & Gillmore, 1997; Marsh & Roche, 2000; Remedios & Lieberman, 2008), so students rated both. The items were "Course difficulty, relative to other courses was" (1 "one of easiest" to 5 "one of most difficult") and "Course workload, relative to other courses was" (1 "very light" to 5 "very heavy"). Also included were class size and instructor sex. Additionally, Griffin (2001) offered evidence that instructor reputation may be associated with student ratings. To assess instructor reputation, students answered this question: "Before taking this course, what did you hear about this instructor?" Reponses ranged from (1 "very bad" to 5 "very good", and 6 "didn't know about the instructor"). For statistical modeling purposes, responses were recoded into one of three categories: negative reputation (score of 1, 2, or 3; about 18.5% of respondents), positive reputation (score of 4 or 5; about 24.8% of respondents), and no reputation (score of 6; about 56.7% of respondents).

As noted earlier there is deliberation about the meaning of the association between student grades, or expected grades, and student ratings (Greenwald & Gillmore, 1997; Marsh & Roche, 1997). Since student anticipated grades predict ratings of instruction, it is important to incorporate expected grades in the analysis of student ratings. To measure expected grades, participants answered two questions. The first sought expected grades with the following wording: "What grade do you think the instructor will assign you in this course?" Responses to this question ranged from 1 (F) to 13 (A+). The second question focused on deserved grade: "What grade do you think you deserve in this course?" Responses to this question were also on the scale of 1 (F) to 13 (A+). Including the second question enables one to calculate an expected grade discrepancy (Griffin, 2004). Grade discrepancy may factor into what Feldman (1997) calls attributional bias and retributional bias. Feldman explains that attributional bias results when one attributes success to one's skills, abilities, or labors, and one attributes failures to causes outside oneself. Retributional bias, notes Feldman, is the tendency for one to reward or punish others for gains or losses. Grade discrepancy could lead to a retribution effect in which students who anticipate receiving grades lower than deserved may attribute this lower grade to instructors and therefore punish them with lower ratings. Holmes (1972) hypothesized, and found support for, a similar retribution effect. To determine grade discrepancy, one may calculate the difference between expected grade and deserved grade (i.e., expected–deserved). There are three possible outcomes with this calculation. First, the difference may be positive which indicates students expect a grade higher than they deserve; a negative difference suggests students anticipate an assigned grade less than deserved; and, finally, no difference suggests both expected and deserved grades are the same so there is no discrepancy. According to retribution theory, a negative discrepancy would result in lower student ratings. The above calculations were performed and results showed that 69.13% ($n = 636$) of sampled students anticipated no discrepancy between expected and deserved grades, 28.04% ($n = 258$) believed their expected grade will be lower than deserved, and only 2.82% ($n = 26$) expected an assigned grade higher than deserved. Given the small number of the latter group, these students were combined with the no discrepancy group. To model grade discrepancy, one dummy variable was created and coded 1 for students who expected a negative grade discrepancy and 0 for those students who anticipated no discrepancy or a positive discrepancy.

In summary, there are 10 outcomes of interest, all measured by student ratings of instruction items, and 12 predictors measured by a variety of items and scales as noted above. Table 1 presents all outcome and predictor variables included in this study.

2.3. Procedure

A proctor administered the instrument during either the penultimate or last week of classes immediately prior to the final examination period during fall and spring semesters. Instructors were required to leave the classroom during administration. The proctor informed students that data from the instrument would not be available to instructors until after the semester ended and course grades assigned. Moreover, the proctor also told students that data would be provided in aggregate form only so individual student responses could not be identified.

3. Results

3.1. Descriptive statistics and correlations

Correlations with the 10 instructional ratings measures were calculated for each of the 12 predictor variables. Both intrinsic motivation and autonomy support were positively and significantly associated with each of the 10 instructional rating measures. Correlations with intrinsic motivation ranged from 0.31 to 0.64 with a mean correlation of about 0.42 across the instructional rating items. For autonomy supportiveness the correlations with the ratings items ranged from 0.23 to 0.42 with a mean correlation

of about 0.33. Based upon the correlations it appears that intrinsic motivation was more strongly associated with instructional ratings than was autonomy supportiveness in instruction. Extrinsic motivation, as measured by student desire for a "good grade" showed little association with student ratings of instruction; the correlations for extrinsic motivation and instructional ratings range from −0.08 to 0.00. Other variables that demonstrated consistent and statistically significant associations with student ratings were the instructor negative reputation dummy variable, negative grade discrepancy, expected grade, grading leniency, and instructor sex.

A limitation with zero-order correlations is that they do not partial the statistical effects of other variables, so regression analyses are needed to obtain more complete assessments of the simultaneous predictive effects of the variables taken together. Two sets of analyses were considered. The first consisted of multilevel models – regression models with random effects – that take into account the natural clustering of data as nested within classes. The second set of analyses examined interactions among the various predictors of interest.

3.2. Mixed model of student ratings

As noted above, the data for this study formed a natural clustering of students grouped within classes, therefore a linear mixed model (Fitzmaurice, Laird, & Ware, 2011), also known as a multilevel model (Snijders & Bosker, 2012) or hierarchal linear model (Raudenbush & Bryk, 2002), was used to examine how each rating dimension regressed on the predictors examined. The following general model was used for each instructional rating dimension:

Student-level.

$$(\text{Student Rating of Instruction Item})_{ij} = \beta_{0j} + \beta_1\,(\text{Intrinsic Motivation})_{ij} + \beta_2\,(\text{Perceived Autonomy})_{ij} + \beta_3\,(\text{Extrinsic Motivation})_{ij} + \beta_4\,(\text{Pos. Reputation dummy})_{ij} + \beta_5\,(\text{Neg. Reputation dummy})_{ij} + \beta_6\,(\text{Neg. Grade Discrepancy})_{ij} + \beta_7\,(\text{Expected Grade})_{ij} + \beta_8\,(\text{Grading Leniency})_{ij} + \beta_9\,(\text{Course Difficulty})_{ij} + \beta_{10}\,(\text{Course Workload})_{ij} + e_{ij}$$

At the class-level, mean ratings of the instructor were modeled with class size and instructor sex:

Class-level.

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\,(\text{Class Size})_j + \gamma_{02}\,(\text{Instructor's Sex})_j + \mu_{0j}$$

Combining the student- and class-level equations yields the following model of instructor ratings:

Combined Model.

$$(\text{Student Rating of Instruction Item})_{ij} = \gamma_{00} + \beta_1\,(\text{Intrinsic Motivation})_{ij} + \beta_2\,(\text{Perceived Autonomy})_{ij} + \beta_3\,(\text{Extrinsic Motivation})_{ij} + \beta_4\,(\text{Pos. Reputation dummy})_{ij} + \beta_5\,(\text{Neg. Reputation dummy})_{ij} + \beta_6\,(\text{Neg. Grade Discrepancy})_{ij} + \beta_7\,(\text{Expected Grade})_{ij} + \beta_8\,(\text{Grading Leniency})_{ij} + \beta_9\,(\text{Course Difficulty})_{ij} + \beta_{10}\,(\text{Course Workload})_{ij} + \gamma_{01}\,(\text{Instructor's Sex})_j + \gamma_{02}\,(\text{Class Size})_j + e_{ij} + \mu_{0j}$$

The combined model was used for each of the 10 instructional rating dimensions identified above. Model estimates, presented in Table 2, were obtained using the "mixed" command, linear models with mixed effects, in Stata 14.1 (StataCorp, 2015). Table 2 contains unstandardized regression estimates.

Regression results showed that both intrinsic motivation and perceived autonomy support were positively and statistically associated with all ten dimensions of instructional ratings. The greater the intrinsic motivation of students, the better ratings provided for instruction. Similarly, the more autonomy support perceived by students, the higher were their ratings of instruction. Since both variables were based upon on the same 5-point scale, it is possible to compare the relative predictive power of these variables. Of the two, intrinsic motivation appears to be the stronger predictor of instructional ratings since its regression coefficients were larger on all ten instructional dimensions. These results are similar to findings reported by Jones (2010) who noted that situational interest, which should be closely related to intrinsic motivation, was the best predictor of student ratings of instruction, and the next best predictors were autonomy support (which Jones labeled as empowerment) and academic caring.

The measure of extrinsic motivation with a focus on course grades demonstrated a statistically significant association with only three instructional dimensions examined: materials well prepared and organized, students shared ideas, and instructor is knowledgeable in the subject of this course. Despite the three statistically significant associations, the coefficient estimates were weak and this suggests that extrinsic motivation with a focus on courses grades plays either no role in ratings, or only a minor and weak role in ratings. The lack of predictive power for extrinsic motivation for course grades is important because it provides some evidence that motivation for grades does not appear to a key factor in determining how students evaluate instruction.

A number of covariates were also included in the regression models. Two of the strongest predictors were dummy variables for negative grade discrepancy and negative instructor reputation. Both variables were statistically significant predictors in eight of the ten models considered, and both displayed negative associations with instructional ratings. The negative grade discrepancy associations show that students who expected a course grade lower than deserved rated lower their instructors, and this finding supports Feldman's (1997) and Holmes' (1972) retribution hypotheses. The negative instructor reputation dummy results indicate that instructors with negative reputations tend to have lower instructor ratings.

Expected grade was also a statistically significant predictor in half of the models examined, and course difficulty was a statistically significant predictor in six of the ten models. Course difficulty demonstrated a positive relationship with instructional ratings, so the greater the perceived difficulty of the course, the higher the average instructional ratings for that course. Expected grade also showed a positive association with ratings—the higher the expected course grade, the better are instructional ratings. As previously noted some argue that the positive relationship between expected grade and instructional ratings could be a signal of a grading leniency effect. To control for this, a separate measure of grading leniency was incorporated into the models and this variable was a significant predictor of instructional ratings in only one of the ten models examined. The one instructional dimension in which grading leniency was a statistically significant predictor was for item asking students to rate whether the instructor used fair and appropriate evaluations of student work. The more lenient the instructor's grading, the higher were ratings of fair and appropriate evaluations of student work. For the other nine instructional dimensions, grading leniency was not a predictor. The remaining variables in the models showed little evidence of association with instructional ratings. These include the positive reputation dummy, course workload, class size, and instructor sex.

3.3. Models of moderation: testing interactions

With any complex regression analysis, some of the predictor variables may form interactions. An interaction means that the nature of the relationship between a predictor and the dependent variable changes, or is moderated, according to levels of another predictor. For example, researchers have speculated that autonomy supportiveness may interact with other motivation variables.

Table 2
Multilevel Regression Results for Student Ratings of Instruction.

	Overall Course		Overall Instructor		Presented Clearly		Materials Organized		Students Shared Ideas	
Fixed Portion of Model	b	se b	b	se b	b	se b	b	se b	b	se b
Student Level										
Intrinsic Motivation	0.56*	0.03	0.34*	0.04	0.45*	0.04	0.38*	0.04	0.21*	0.03
Perceived Autonomy	0.11*	0.03	0.17*	0.03	0.10*	0.03	0.16*	0.03	0.17*	0.03
Extrinsic-Grade Motivation	0.01	0.02	0.01	0.02	0.04	0.02	0.07*	0.02	0.06*	0.02
Instructor Reputation										
Pos. Reputation	0.03	0.06	0.12	0.06	0.01	0.07	0.02	0.07	0.04	0.06
Neg. Reputation	−0.27*	0.08	−0.38*	0.08	−0.05	0.09	−0.12	0.08	−0.21*	0.07
Neg. Grade Discrepancy	−0.21*	0.06	−0.27*	0.06	−0.21*	0.07	−0.08	0.06	−0.15*	0.06
Expected Grade	0.11	0.04	0.11	0.05	0.16*	0.05	0.09	0.05	0.08	0.04
Grading Leniency	0.02	0.02	0.04	0.03	0.04	0.03	0.02	0.03	0.04	0.02
Course Difficulty	0.10*	0.03	0.13*	0.04	0.08	0.04	0.06	0.04	0.10*	0.03
Course Workload	0.01	0.03	0.01	0.04	−0.05	0.04	0.04	0.04	−0.01	0.03
Class Level										
Class Size	−0.01	0.01	−0.01	0.01	−0.01	0.01	−0.01	0.01	0.01	0.01
Instructor's Sex	−0.29	0.14	−0.36	0.18	−0.27	0.17	−0.22	0.13	0.03	0.09
Model Intercept	3.78*	0.10	4.21*	0.12	4.12*	0.12	4.29*	0.09	4.60*	0.06
Random Portion of Model										
Class-level variance	.19*		.32*		.29*		.17*		.06*	
Student-level variance	.46*		.54*		.59*		.57*		.43*	
R^2 (total variance modeled)	0.47		0.37		0.31		0.28		0.24	

	Students Could Seek Help		Fair Evaluation of Students		Concern for Students		Feedback Helpful		Instructor Knowledgeable	
Fixed Portion of Model	b	se b	b	se b	b	se b	b	se b	b	se b
Student Level										
Intrinsic Motivation	0.26*	0.04	0.28*	0.03	0.27*	0.03	0.31*	0.03	0.29*	0.03
Perceived Autonomy	0.24*	0.03	0.16*	03	0.22*	0.03	0.20*	0.03	0.10*	0.02
Extrinsic-Grade Motivation	0.02	0.02	0.03	0.02	0.01	0.02	0.03	0.02	0.06*	0.02
Instructor Reputation										
Pos. Reputation	0.04	0.06	−0.05	0.06	0.04	0.06	−0.02	0.06	−0.01	0.06
Neg. Reputation	−0.28*	0.08	−0.45*	0.08	−0.26*	0.08	−0.24*	0.08	−0.21*	0.07
Neg. Grade Discrepancy	−0.23*	0.06	−0.29*	0.06	−0.21*	0.06	−0.17*	0.06	−0.05	0.05
Expected Grade	0.13*	0.05	0.14*	0.05	0.11*	0.04	0.15*	0.05	0.03	0.04
Grading Leniency	0.04	0.03	0.13*	0.02	0.05	0.02	0.07	0.03	0.03	0.02
Course Difficulty	0.09	0.04	0.10*	0.03	0.09*	0.03	0.07	0.04	0.12*	0.03
Course Workload	−0.01	0.04	0.03	0.03	0.03	0.03	.08*	0.04	0.02	0.03
Class Level										
Class Size	0.01	0.01	0.01	0.01	−0.01	0.01	−0.01	0.01	−0.01	0.01
Instructor's Sex	−0.16	0.11	−0.02	0.11	−0.11	0.12	−0.16	0.12	−0.07	0.09
Model Intercept	4.49*	0.08	4.43*	0.08	4.46*	0.08	4.35*	0.09	4.58*	0.06
Random Portion of Model										
Class-level variance	0.09*		0.11*		0.12*		0.14*		0.07*	
Student-level variance	0.55*		0.50*		0.46*		0.52*		0.41*	
R^2 (total variance modeled)	0.32		0.38		0.35		0.32		0.25	

Note: Negative Grade Discrepancy coded 1 if grade lower than believed deserved, 0 otherwise; Positive Reputation dummy coded 1 if student rated instructor as having positive reputation, 0 otherwise; and Negative Reputation dummy coded 1 if student rated instructor as having negative reputation, 0 otherwise. Instructor Sex coded 1 for males and 0 for females. All variables centered at their means except for dummy variables (Grade Discrepancy, Reputation, and Instructor Sex). b = unstandardized regression coefficients. n = 914 students in 47 courses. *$p < 0.01$.

Vansteenkiste et al. (2004) reported an interaction between autonomy support and goal orientation (intrinsic vs. extrinsic) when modeling student performance. They found that a controlling environment tended to suppress student performance for both those with intrinsic and extrinsic goal orientations, and that an autonomy-supportive environment tended to increase the difference in performance between those with intrinsic and extrinsic goal orientations. Similarly, when modeling college students' final course grades, Black and Deci (2000) reported a moderating effect between instructor autonomy support and student intrinsic or extrinsic motivating reasons for taking the course. Black and Deci found a positive relation between teacher autonomy supportiveness and student final course grades for students with extrinsic reasons for enrolling in a course, but for students who entered with course with an intrinsic interest, there was no statistical evidence that instructor autonomy supportiveness predicted final course grades.

To learn whether autonomy support, intrinsic motivation, and extrinsic motivation interacted to form moderated relations with student ratings of instruction, three interactions were tested for each of the ten student rating dimensions: autonomy support × intrinsic motivation, autonomy support × extrinsic motivation, and intrinsic motivation × extrinsic motivation. Table 3 presents regression coefficient estimates of these interactions for the ten student ratings items. Normally one would also present the main effect coefficient estimates, but coefficient estimates for models

Table 3
Multilevel Regression Results for Interactions Among Autonomy Support, Intrinsic Motivation, and Extrinsic Motivation.

	Overall Course		Overall Instructor		Presented Clearly		Materials Organized		Students Shared Ideas	
Interaction Term	b	se b	b	se b	b	se b	b	se b	b	se b
IM × AS	0.02	0.03	−0.08*	0.03	−0.15*	0.03	−0.14*	0.03	−0.17*	0.02
IM × EM	0.01	0.02	0.03	0.02	0.01	0.02	−0.04	0.02	−0.04	0.02
AS × EM	0.02	0.02	0.01	0.02	−0.05	0.02	−0.02	0.02	−0.06*	0.02

	Students Could Seek Help		Fair Evaluation of Students		Concern for Students		Feedback Helpful		Instructor Knowledgeable	
Interaction Term	b	se b	b	se b	b	se b	b	se b	b	se b
IM × AS	−0.19*	0.03	−0.14*	0.03	−0.17*	0.02	−0.17*	0.03	−0.11*	0.02
IM × EM	−0.01	0.02	−0.02	0.02	0.02	0.02	0.04	0.02	−0.05	0.02
AS × EM	−0.06*	0.02	−0.03	0.02	−0.08*	0.02	−0.04	0.02	−0.03	0.02

Note: IM = Intrinsic Motivation, AS = Autonomy Support, and EM = Extrinsic Motivation. Tables includes only interaction coefficients; to see coefficients for other predictors in mode, see Table 2. Note that values between models in Table 2 and models with interactions were virtually unchanged, so only interaction coefficients reported here. Three-way interactions among IM, AS, and EM were tested and none statistically significant at the 0.01 level. b = unstandardized regression coefficients. n = 914 students in 47 courses.

* $p < 0.01$.

with interactions were mostly unchanged from those presented in Table 2, so Table 3 presents only the unstandardized interaction estimates.

The results for Table 3 show that at least one interaction was statistically significant for nine of the ten student ratings items. The rating item for the overall course was the only item to have no significant interactions. In the remaining nine instructional ratings items, the interaction between autonomy support and intrinsic motivation was statistically significant at the 0.01 level. In each of these cases, the estimated interaction coefficient was similar across regression models and ranged from −0.08 to −0.19 with a mean estimate of −0.14. Only three of the ten instructional ratings items produced statistically significant interactions between autonomy support and extrinsic motivation, and these three represented small statistical effects (mean coefficient estimate was b = −0.07). There were no statistically significant interactions between intrinsic and extrinsic motivation.

The key finding from the analyses presented in Table 3 is that there is strong evidence of an interaction between autonomy support and intrinsic motivation when predicting student ratings of instruction. As noted, the mean of the nine significant interactions was −0.14. This is the same value as the interaction reported for the materials well prepared and organized rating item (henceforth referred to as the materials organization item), so this item will be used to illustrate how the interaction between autonomy support and intrinsic motivation can be interpreted for each of the nine rating items for which the interaction was significant. Table 4 presents results for the moderation effect between intrinsic motivation and autonomy support for the materials organization item.

Table 4 reports the expected regression slope for both intrinsic motivation and autonomy support for three different levels of each variable. For intrinsic motivation, Table 4 presents three coefficient estimates: one for students who judge autonomy support to be high (i.e., one standard deviation above the mean for autonomy support), one for students who judge autonomy support to be at the mean value, and one for students who judge autonomy support to be low (i.e., one standard deviation below the mean for autonomy support). The respective slope values for intrinsic motivation are 0.24, 0.39, and 0.53. These coefficients show that intrinsic motivation positively, and significantly, predict ratings on the materials organization item no matter what level of autonomy support provided in a class. However, the estimates also show that intrinsic motivation becomes a weaker predictor of ratings as instructor autonomy supportiveness increases. The slope drops from 0.53 to 0.24 as autonomy supportiveness increases from one standard deviation below to one standard deviation above the mean of autonomy supportiveness.

In terms of autonomy supportiveness, Table 4 shows that student intrinsic motivation plays an important role in determining whether autonomy supportiveness predicts student ratings for the materials organization item. For students with high levels of intrinsic motivation (i.e., one standard deviation above the mean of intrinsic motivation), autonomy support does not predict ratings on the materials organization item (b = 0.05, p = 0.19), but as intrinsic motivation declines, autonomy support becomes more important in predicting ratings for this item. For students with low levels of intrinsic motivation (i.e., one standard deviation below the mean for intrinsic motivation), the estimated slope for autonomy support predicting ratings of materials organization is b = 0.28 ($p < 0.01$). A graphical display may help clarify these moderated effects.

Fig. 1 shows the nature of the relationship between autonomy support and student ratings on the materials organization item for different levels of intrinsic motivation. The relationships shown in Fig. 1 are similar to those found for other student rating items, except for the overall course item for which no interactions were detected. Fig. 1 reveals that when intrinsic motivation is high, the slope between autonomy support and ratings on the materials organization item is nearly flat which indicates little predictive

Table 4
Interactive Statistical Effects for Intrinsic Motivation and Autonomy Support Predicting Student Ratings of Course Materials Well Prepared and Organized.

	Intrinsic Motivation			Autonomy Support	
	b	se		b	se
High Auto. Support (+1 SD)	0.24*	0.05	High Intrinsic Mot. (+1 SD)	0.05	0.04
Auto. Support at Mean	0.39*	0.04	Intrinsic Mot. at Mean	0.17*	0.03
Low Auto. Support (−1 SD)	0.53*	0.05	Low Intrinsic Mot. (−1 SD)	0.28*	0.04

Note: The first set of coefficients reported in this table represent the relationship between intrinsic motivation and course materials well prepared and organized student rating item for three levels of autonomy support. The second set of coefficients show the relationship between autonomy support and student ratings of course material preparation and organization for three levels of intrinsic motivation. b = unstandardized regression coefficients.

* $p < 0.01$.

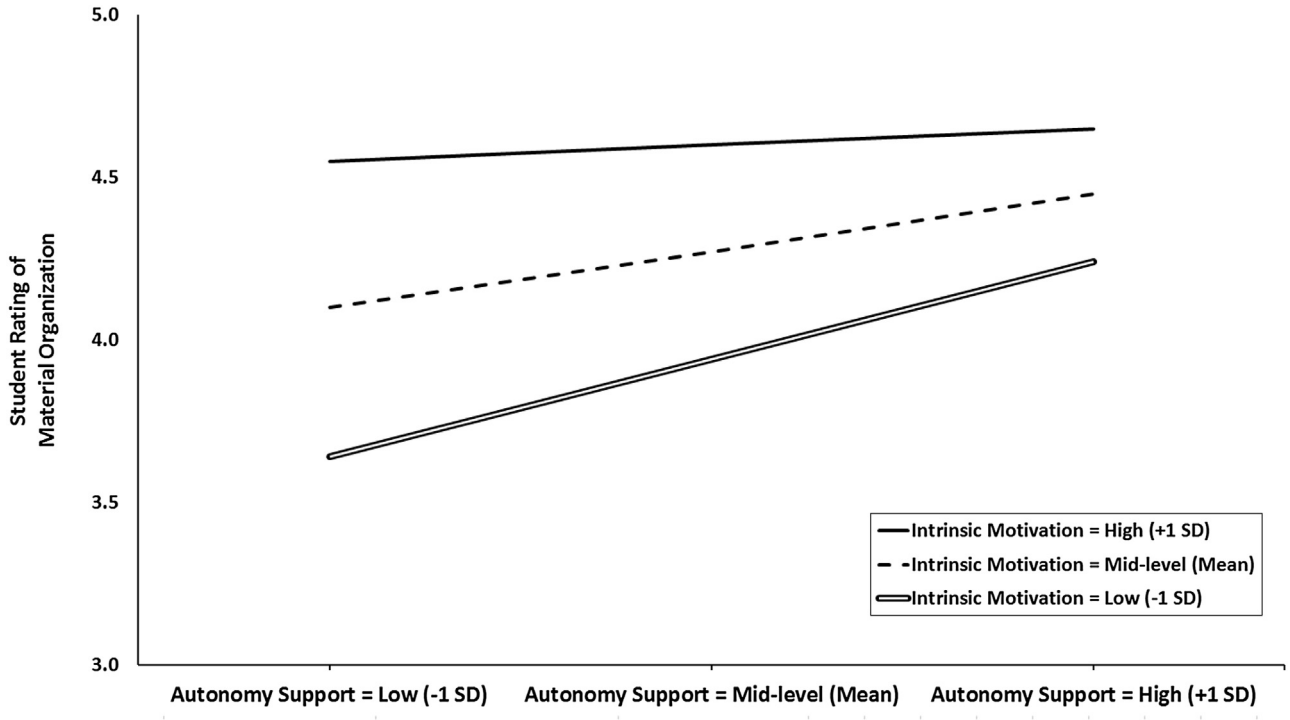

Fig. 1. Interaction between Autonomy Support and Intrinsic Motivation for Student Ratings of Materials Well Prepared and Organized Dimension.

power for autonomy support. When intrinsic motivation is near the mean or lower, the relationship between autonomy support and ratings for the materials organization item becomes stronger (i.e., steeper) and this suggests autonomy support becomes a more effective predictor of student ratings. The interaction results displayed in Fig. 1 also shows a ratings gap among the three levels of intrinsic motivation when autonomy support is low; however, when autonomy support is high, the ratings gap closes among the three levels of intrinsic motivation. This moderation effect suggests that instructors of courses that engender lower intrinsic motivation among students can close the potential ratings gap, or simply increase their student ratings, by offering greater autonomy supportiveness within the classroom.

4. Discussion

The results of this study confirm expectations and previous research by showing that instructor autonomy supportiveness (Filak & Sheldon, 2003; Filak & Sheldon, 2008; Jones, 2010) and student intrinsic motivation (Olivares, 2001) are both positively associated with student ratings of instruction. Further, the findings of this study expand the literature on student ratings by demonstrating that both instructor autonomy supportiveness and intrinsic motivation are associated with multiple dimensions of student ratings of instruction. Additionally, while previous studies have shown autonomy supportiveness to form interactions with other predictors – for example with goal orientation (intrinsic/extrinsic) to predict graded performance (Vansteenkiste et al., 2004), with structure in teaching to model student self-regulated learning (Sierens, Vansteenkiste, Goossens, Soenens, & Dochy, 2009), or with intrinsic/extrinsic motivating reasons for taking a course to predict final course grades (Black & Deci, 2000) – the current study showed that autonomy supportiveness moderates, and is moderated by, intrinsic motivation to predict multiple dimensions of student ratings of instruction.

How might these findings be relevant to instructors? Research shows that teachers can learn to incorporate within their classes autonomy-supportive behaviors, and this can enhance both student learning and motivation (Deci et al., 1991; Reeve & Cheon, 2016). The results of the current study show that autonomy-supportive behavior becomes a stronger predictor of instructional ratings the lower a student's level of intrinsic motivation for a course or its content. If it is possible to infer from ratings about teaching effectiveness and instructional behavior, then these findings suggest that instructors of courses of low inherent interest or with lowly motivated students would do well to adopt more autonomy-supportive instructional behavior within their classrooms.

As a practical suggestion, Stefanou et al. (2004), Assor et al. (2002), Reeve (2015), and Reeve and Cheon (2016) note that autonomy support can take a number of forms, and some of the more common include giving voice to students so they can express their thoughts, providing students with options among educational activities, removing or minimizing controlling behavior over students, and finding ways to encourage students to take charge over their own learning. Assor et al. explain that autonomy support is not about relinquishing control and responsibility within the classroom, but instead autonomy support provides a means to help students generate and seek self-relevant, course-related objectives within a class.

There are myriad approaches to providing autonomy support within one's class. For example, one approach to recognizing students' perspectives and giving them voice is to seek and value their input. Chen and Hoshower (2003), for instance, found that university students want to offer input about instruction if they believe that input will be valued. Chen and Hoshower recommended that instructors include an example of how they used student ratings of instruction to adjust their teaching for the better. Providing such an example shows students that their feedback is important and plays a role in shaping instructional activities. Griffin and Pool (1998) documented a comparable procedure for soliciting student feedback multiple times throughout a semester. Griffin and Pool reported that this process provided rich information for instructional improvement and resulted in a more

positive learning environment. Stead (2005) reported similar results in his review of the "one-minute paper," another technique for soliciting student feedback at the end of class sessions. Additionally, Reeve and Cheon (2016) describe a similar method for collecting student suggestions at the end of a lesson and incorporating those into future lessons.

In summary, results of this study should be encouraging to instructors because these findings indicate that students recognize instructor behaviors designed to foster a positive learning environment, and the results show that when instructors incorporate autonomy-supportive behaviors into a college class, students provide correspondingly higher ratings. Reeve et al. (2004) research demonstrates that instructors can learn to employ within their classes autonomy-supportive behaviors. See Reeve (2009, 2015), Reeve and Cheon (2016), and Stefanou et al. (2004) for detailed, concrete examples of how one may operationalize autonomy-supportive instruction within one's class.

References

Abrami, P. C., d'Apollonia, S., & Rosenfield, S. (2007). The dimensionality of student ratings of instruction: What we know and what we do not. In R. P. Perry, & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 385–456).New York: Springer.

Apodaca, P., & Grad, H. (2005). The dimensionality of student ratings of teaching: Integration of uni- and multidimensional mode. *Studies in Higher Education*, *30*, 723–748.

Assor, A., Kaplan, H., & Roth, G. (2002). Choice is good, but relevance is excellent: Autonomy-enhancing and suppressing teacher behaviours predicting students' engagement in schoolwork. *British Journal of Educational Psychology*, *72*, 261–278.

Benton, S. L., & Cashin, W. E. (2012). *Student ratings of teaching: A Summary of research and literature. (IDEA Paper #50).* Manhattan, Kansas, USA: The IDEA Center.

Bieg, S., Backes, S., & Mittag, W. (2011). The role of intrinsic motivation for teaching, teachers' care and autonomy support in students' self-determined motivation. *Journal of Educational Research Online*, *1*, 122–140.

Black, A. E., & Deci, E. L. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education*, *84*, 740–756.

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.

Cashin, W. E., & Downey, R. G. (1992). Using global student rating items for summative evaluation. *Journal of Educational Psychology*, *84*, 563–572.

Chen, Y., & Hoshower, L. B. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment & Evaluation in Higher Education*, *28*, 71–88.

Chen, G., & Watkins, D. (2010). Stability and correlates of student evaluations of teaching at a Chinese university. *Assessment & Evaluation in Higher Education*, *35*, 675–685.

Church, M. A., Elliot, A. J., & Gable, S. L. (2001). Perceptions of classroom environment, achievement, goals, and achievement outcomes. *Journal of Educational Psychology*, *93*, 43–54.

Covington, M. V. (2000). Goal theory, motivation: and school achievement: An integrative review. *Annual Review of Psychology*, *51*, 171–200.

Darwin, S. (2010). Exploring critical conceptions of student-led evaluation in Australian higher education. *Research and Development in Higher Education*, *33*, 203–212.

Deci, E., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determiantion perspective. *Educational Psychologist*, *26*, 325–346.

Deci, E. L., Nezlek, J., & Sheinman, L. (1981). Characteristics of the rewarder and intrinsic motivation of the rewardee. *Journal of Personality and Social Psychology*, *40*(1), 1–10.

Deci, E. L., Schwartz, A. J., Sheinman, L., & Ryan, R. M. (1981). An instrument to assess adults' orientations toward control versus autonomy with children: Reflections on intrinsic motivation and perceived competence. *Journal of Educational Psychology*, *73*(5), 642–650.

Feldman, K. A. (1997). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry, & J. C. Smart (Eds.), *Effective teaching in higher education: Research and practice* (pp. 368–395).New York: Agathon.

Feldman, K. A. (1998). Reflections on the study of effective college teaching and student ratings: One continuing question and two unresolved issues. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (pp. 35–74).New York: Agathon.

Filak, V. F., & Sheldon, K. M. (2003). Student psychological need satisfaction and college teacher-course evaluations. *Educational Psychology*, *23*, 235–247.

Filak, V. F., & Sheldon, K. M. (2008). Teacher support, student motivation, student need satisfaction, and college teacher course evaluations: Testing a sequential path model. *Educational Psychology*, *28*(6), 711–724.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis*, 2nd ed. Hoboken, N.J: Wiley.

Garcia, T., & Pintrich, P. R. (1996). The effects of autonomy on motivation and performance in the college classroom. *Contemporary Educational Psychology*, *21*, 477–486.

Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, *52*, 1209–1217.

Greimel-Fuhrmann, B., & Geyer, A. (2003). Students' evaluation of teachers and instructional quality—Analysis of relevant factors based on empirical evaluation research. *Assessment & Evaluation in Higher Education*, *28*, 229–238.

Griffin, B. W., & Pool, H. (1998). Monitoring and improving instructional practices (and are student evaluations valid?). *Journal of Research and Development in Education*, *32*, 1–9.

Griffin, B. W. (2001). Instructor reputation and student ratings of instruction. *Contemporary Educational Psychology*, *26*, 534–552.

Griffin, B. W. (2004). Grading leniency, grade discrepancy, and student ratings of instruction. *Contemporary Educational Psychology*, *29*, 410–425.

Haerens, L., Aelterman, N., Vansteenkiste, M., Soenens, B., & Van Petegem, S. (2015). Do perceived autonomy-supportive and controlling teaching relate to physical education students" motivation experiences through unique pathways? Distinguishing between the bright and dark side of motivation. *Psychology of Sport and Exercise*, *16*, 26–36.

Hendry, G. D., & Dean, S. J. (2002). Accountability, evaluation of teaching and expertise in higher education. *The International Journal of Academic Development*, *7*, 75–82.

Holmes, D. S. (1972). Effects of grades and disconfirmed grade expectations on students' evaluations of their instructor. *Journal of Educational Psychology*, *63*, 130–133.

Howard, G. S., & Maxwell, S. E. (1980). Correlation between student satisfaction and grades: A case of mistaken causation? *Journal of Educational Psychology*, *72*, 810–820.

Hoyt, D., & Lee, E. (2002). *Basic data for the revised IDEA system. (IDEA technical report No. 12).* Manhattan, Kansas, USA: The IDEA Center.

Husbands, C., & Fosh, P. (1993). Student's evaluation of teaching in higher education: Experiences from four European countries and some implication of the practice. *Assessment & Evaluation on Higher Education*, *18*, 95–114.

Husbands, C. T. (1998). Implications for the assessment of the teaching competence of staff in higher education of some correlates of students' evaluations of different teaching styles. *Assessment & Evaluation in Higher Education*, *23*, 117–139.

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.

Leckey, J., & Neill, N. (2001). Quantifying quality: The importance of student feedback. *Quality in Higher Education*, *7*, 19–32.

Lepper, M. R., Corpus, J. H., & Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of Educational Psychology*, *97*(2), 184–196.

Linnenbrink, E. A., & Pintrich, P. R. (2002). Motivation as an enabler for academic success. *School Psychology Review*, *31*, 313–327.

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*, 1187–1197.

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*, 202–228.

Marsh, H. W. (1980). The influence of student, course, and instructor characteristics on evaluations of university teaching. *American Educational Research Journal*, *17*, 219–237.

Marsh, H. W. (1983). Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics. *Journal of Educational Psychology*, *75*, 150–166.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, *11*, 253–388.

Murray, H. G. (1997). Effective teaching behaviors in the college classroom. In R. P. Perry, & J. C. Smart (Eds.), *Effective teaching in higher education: Research and practice* (pp. 171–204).New York: Agathon.

Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. *Theory and Research in Education*, *7*(2), 133–144.

Olivares, O. J. (2001). Student interest, grading leniency, and teacher ratings: A conceptual analysis. *Contemporary Educational Psychology*, *26*, 382–399.

Pintrich, P. R., Smith, D., Garcia, T., & McKeachie, W. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).* Ann Arbor, Michigan: University of Michigan.

Prave, R. S., & Baril, G. L. (1993). Instructor ratings: Controlling for bias from initial student interest. *Journal of Education for Business*, *68*, 362–366.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models*, 2nd ed. Thousand Oaks: Sage Publications.

Reeve, J., & Cheon, S. H. (2016). Teachers become more autonomy supportive after they believe it is easy to do. *Psychology of Sport and Exercise*, *22*, 178–189.

Reeve, J., Jang, H., Carrell, D., Jeon, S., & Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion*, *28*, 147–169.
Reeve, J. (2009). Why teachers adopt a controlling motivating style toward students and how they can become more autonomy supportive. *Educational Psychologist*, *44*(3), 159–175.
Reeve, J. (2015). Giving and summoning autonomy support in hierarchical relationships. *Social and Personality Psychology Compass*, *9/8*, 406–418.
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.
Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*, 54–67.
Saroyan, A., & Amundsen, C. (2001). Evaluating university teaching: Time to take stock. *Assessment & Evaluation in Higher Education*, *26*, 341–353.
Sierens, E., Vansteenkiste, M., Goossens, L., Soenens, B., & Dochy, F. (2009). The synergistic relationship of perceived autonomy support and structure in the prediction of self-regulated learning. *British Journal of Educational Psychology*, *79*, 57–68.
Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*, 2nd ed. Los Angeles: Sage.
StataCorp (2015). *Stata multilevel mixed-effects reference Manual: Release 14*. College Station, TX: StataCorp LP.
Stead, D. (2005). A review of the one-minute paper. *Active Learning in Higher Education*, *6*, 118–131.
Stefanou, C. R., Perencevich, K. C., DiCintio, M., & Turner, J. C. (2004). Supporting autonomy in the classroom: Ways teachers encourage decision making and ownership. *Educational Psychologist*, *39*, 97–110.
Stipek, D. (1998). *Motivation to learn: From theory to practice*, 3rd ed. Boston: Allyn and Bacon.
Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: The synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology*, *87*, 246–260.
Williams, G. C., & Deci, E. L. (1996). Internalization of biopsychosocial values by medical students: A test of self-determination theory. *Journal of Personality and Social Psychology*, *70*(4), 767–779.