International Journal of Teaching and Learning in Higher Education 2010, Volume 22, Number 3, 331-336
http://www.isetl.org/ijtlhe/ ISSN 1812-9129

Course Grades, Quality of Student Engagement, and Students' Evaluation of Instructor

Steven Culver
Virginia Tech

Students' evaluation of faculty and courses continue to be the most often used gauge in higher education of how well courses are taught. Faculty are particularly concerned that student ratings are highly associated with the grades students expect to receive. However, newer research on student engagement suggests that it is students' own interaction with the course material that determines their evaluation of the course. The purpose of this study then was to examine (1) whether the grades students expected in the course affected the overall evaluation of the instructor, (2) whether the students' quality of engagement in the course affected the overall evaluation of the instructor, and (3) whether students' quality of engagement moderates the relationship between expected grades and overall evaluation of the instructor. Results indicate that students' engagement with the course material significantly moderates the relationship between expected grades and overall rating of instructor.

Students' evaluation of faculty and courses continue to be the most often used gauge in higher education of how well courses are taught, despite questions regarding their validity. In the last decade, Seldin (1999) noted the predominance of the student evaluation system. Since the early 1970s, a great deal of attention has been paid to research on student ratings of instruction (Spooren, Mortelmans, & Denekins, 2007) and indeed, there were well over 2000 studies on the topic referenced in the ERIC system even five years ago (Centra, 2003). Specifically, much of the research and debate centers on the validity of these student ratings. Though the majority of these studies tend to conclude that these evaluations are reliable and valid when compared to other measures of effective teaching (Centra, 2003), there are also studies indicating that ratings are biased by such factors as workload (Marsh, 2001), student effort (Centra & Gaubatz, 2000), and grading leniency (Griffin, 2004), Student ratings have also been found to be related to students' sense of involvement in the course (Remedios & Lieberman, 2008).

Of particular concern to faculty is the perceived relationship between grades and student evaluations. Many faculty believe that they are, at least until they are tenured, held hostage by students because they believe that lower student grades will result in lower course evaluations, a key element in their faculty evaluation process related to tenure and promotion. This belief contributes to doubts about the validity of students' perceptions of the overall performance of an instructor (Sproule, 2002), especially since students are not typically educated about the importance and use of these ratings (Theall & Franklin, 2001). As Knapper (2001) has succinctly pointed out, "it is a rare campus where [student ratings of university teachers] are accepted with equanimity" (p. 3). Consequently, Eiszler (2002) notes, that despite the many studies on student evaluations, the question still remains regarding the relationship between grading leniency and overall ratings.

Another influence on student perceptions of their classroom experience relates to how difficult they perceive the course to be and, what some have labeled, course workload. Factors typically measured that defined this concept include hours per week spent studying (Gillmore & Greenwald, 1994; Greenwald & Gillmore, 1997) or a more general measure of course difficulty (Marsh & Roche, 2000; Centra & Gaubatz, 2000). Broad measures of course difficulty or workload could, however, be problematic. Centra (2003) suggests that hours spent on coursework, for instance, should be refined by dividing those hours into "good" hours (deemed valuable by students) and "bad" hours, a distinction documented by Marsh (2001). Students' engagement with the material and the class is described more accurately by the "good" hours than the "bad."

Student engagement is a broad construct recognized as providing information to measure students' involvement with their learning (Shulman, 2002), an indirect measure of educational outcomes (Ewell & Jones, 1996), and a measure of students' interaction with their universities (Kuh, et al., 2005). As Coates (2005) has described the process, "learning is influenced by how an individual participates in educationally purposeful activities" (p. 26). Students who are more engaged in their educational processes are more likely to be active and collaborative learners (Pascarella & Terenzini, 2005). Thus, spending a lot of hours outside of class studying or doing lab work is not necessarily a measure of engagement. Rather, this

time spent would only contribute to the engagement of students if they felt that the time spent was worthwhile. So the time and effort required for a class, coupled with a student's perception of the educational value of out-of-class assignments, would present a proxy measure of not only the time spent on the coursework but also a measure of the quality of the engagement with the material.

The purpose of this study then was to examine (1) whether the grades students expected in the course affected the overall evaluation of the instructor, (2) whether the students' quality of engagement in the course affected the overall evaluation of the instructor, and (3) whether students' quality of engagement moderates the relationship between expected grades and overall evaluation of the instructor.

Table 1
Percentage of Responses from Undergraduate Students (N=320,557) to Study Items on Course Evaluation Forms

Overall Rating of this Instructor	
Poor	2.2%
Fair	7.0%
Good	32.2%
Excellent	59.0%
Educational Value of Out-of-Class Assignments	
Poor	2.6%
Fair	12.4%
Good	42.0%
Excellent	35.4%
Time and Effort Required	
Less than Average	10.3%
Average	62.4%
More than Average	26.9%
The Grade I Expect in this Course	
A	40.9%
B	40.6%
C	11.5%
D	1.2%
F	0.1%
My Academic Level	
Freshman	26.6%
Sophomore	26.2%
Junior	22.5%
Senior	24.7%
I Would Rate my Gains in this Course Compared with Similar Courses as Follows	
Knowledge of principles theories…	
Less than Average	6.2%
Average	59.6%
More than Average	34.2%
Logical thinking and problem solving ability…	
Less than Average	10.2%
Average	65.2%
More than Average	24.6%
Appreciation of subject matter and discipline…	
Less than Average	7.3%
Average	55.2%
More than Average	37.5%

Method

Between the Fall of 2002 and the Spring of 2007, students at a Research I, state-supported university in the southeastern United States submitted 350,846 course evaluations. The course evaluation form is completed anonymously (with no student identifiers) by students in each course section near the end of the semester. Collected paper forms are then forwarded to a central administrative office for processing and generating reports to individual faculty, department chairpersons, and deans.

The form includes sixteen questions divided into three sections: instructor ratings, course ratings, and course descriptors. In the instructor ratings section, students are asked to rate, on a four-point (poor, fair, good, excellent) Likert-scale, six individual characteristics of the instructor as well as an "overall rating of instructor." The six individual characteristics include such items as "apparent knowledge of subject matter," "success in communicating or explaining subject matter," "degree to which subject matter was made stimulating or relevant," "concern and respect for students as individuals," "fairness in assigning grades," and "administration of the class and organization of materials." There are three items in the Course Ratings section, with "adequacy of textbook and other study materials" and "educational value of out-of-class assignments" using the same four-point rating scale as the previous seven items. The third item in this section, "Time and effort required," requires students to respond with one of three choices: "less than average," "average," or "more than average." The Course Descriptor section contains items asking students to identify whether or not the course was a requirement for their major or an elective, to indicate their academic level (freshman, sophomore, junior, senior, master's, doctoral), to indicate "the grade I expect in this course (F, D, C, B, A), and to indicate level (less than average, average, more than average) of gains related to knowledge of principles and theories, logical thinking, and appreciation of the subject matter. Percentages of responses for each category for variables included in this study are shown in Table 1. For the purposes of this study, only those evaluations completed by students indicating they were undergraduates (freshman, sophomore, junior, or senior) were analyzed.

The Dependent Variable

The dependent variable of this study was student responses on a Likert-scale of 1-4 (poor, fair, good, excellent) to the item—"overall rating of this instructor." Other items on the instrument solicit opinions regarding aspect of instructor performance, such as apparent knowledge of subject matter, success in communicating or explaining subject matter, or concern and respect for students as individuals. However, more weight is typically placed on the "overall rating" by tenure and

Table 2
Analysis of Variance for Overall Evaluation of Instructor

Source	df	F	p	Partial eta squared
Quality of Engagement	2	6625.20	.000	.047
Expected Grade	3	2769.16	.000	.003
Quality of Engagement x Expected Grade	6	152.29	.000	.003
Error	269244	(.384)		

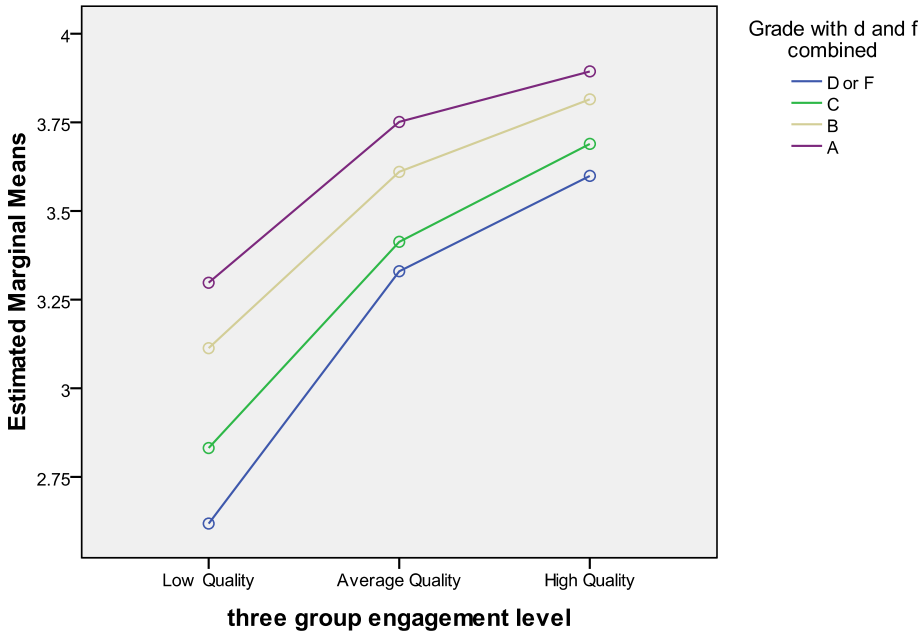
R squared = .24

Table 3
Average Overall Rating of Instructor by Grade Expected by Level of Quality Engagement

	Grade Expected				
	F/D	C	B	A	Marginals
Low Quality	2.62	2.83	3.11	3.30	2.97
	(.012)	(.005)	(.003)	(.003)	(.003)
Average Quality	3.33	3.41	3.61	3.75	3.53
	(.019)	(.006)	(.003)	(.003)	(.005)
High Quality	3.60	3.69	3.82	3.89	3.73
	(.031)	(.008)	(.004)	(.003)	(.008)
Total	3.18	3.31	3.51	3.65	3.41
	(.013)	(.004)	(.002)	(.002)	(.003)

Note: Standard errors are shown in parentheses.

Figure 1
Estimated Marginal Means of Overall Rating of this Instructor



promotion committees, and it is this item that becomes of most concern to instructors.

Independent Variables

The independent variables in this study include the students' expected course grade (as in Centra, 2003) as measured by their response to the item: "The grade I expect to receive in this course is . . . ." Response choices were F, D, C, B, A. Classes where students were graded only on a pass or fail scale (P/F) were removed from the data base prior to analyses, as well as those students who were taking a graded course P/F. In addition, no differences were found between those

expecting F's and D's and their correlation with the dependent variable. Consequently, the five grade groups were reduced to four: F/D, C, B, A.

The second independent variable was Quality of Engagement as measured by students' responses to five items on the course evaluation form. The first item was "Educational value of out of class assignments," to which students could respond using a four-point Likert scale – poor, fair, good, excellent. The second item was "Time and effort required." Students responded to this item using a three-point scale: less than average, average, more than average. The other three items used to create the Quality of Engagement scale were items related to students' perceptions of gains in the course. These gains focused on the areas of "knowledge of principals, theories," "enhanced critical thinking," and "appreciation for the subject matter/field." For each of these items, students were asked to respond with one of three choices to provide their perceptions of this class as compared to other courses they had taken at the university: (1) below average, (2) average, or (3) above average. Students' responses were summed for these five items, creating a scale ranging from a low score of 5 to a highest possible score of 16, with an overall mean of 12.11 and a standard deviation of 2.12. Alpha reliability for this scale was. 72. Based on their scale scores, students were then divided into three groups according to their engagement in the class: Low quality of engagement, Average quality of engagement, and High quality of engagement.

To address the three questions guiding this study – whether expected grades affect overall evaluation of instructor, whether students' engagement affects overall evaluation of instructor, and whether students' engagement moderates the relationship between expected grades and overall evaluation of instructor – a two-way analysis of variance (ANOVA) was conducted.

Results

Table 2 shows the results of the two-way (3 x 4) between-groups analysis of variance conducted to explore the impact of expected student grade and quality of engagement on the overall evaluation of the instructor of the course. Though main effects for Expected Grades [$F(3, 3061.28) = 1020.43$, $p<.01$] and Quality of Engagement [$F(2, 4882.72) = 2441.36$, $p<.01$] were both statistically significant, the interaction effect was significant [$F(6, 336.70) = 56.12$, $p<.01$], indicating that the relationship between the overall rating given the instructor and the student's expected grade is moderated by the student's quality of engagement. In other words, both variables are necessary to predict the Overall Evaluation of Instructor. The cell means and marginal means demonstrating this interaction are presented in Table 3 and the graphic depiction of the interaction is show in Figure 1. As shown in both Table 3 and Figure 1, for example, students who believe they will receive a D or F in the course, but who are also heavily engaged in the course, provide an overall rating of instructor that is higher than students who believe they will receive an A or B but are in the lowest engagement group. The highly engaged D/F students also rate their instructors more highly than the C students who are in the lowest and the average engagement groups.

Conclusions

Despite faculty concerns that students rate faculty more highly when they expect higher grades in the course, the results of these analyses demonstrate that this relationship is moderated significantly by the quality of engagement of the student in that course. With these data, one would be more likely to conclude that engaging students in quality efforts in a course, rather than giving them high grades, would increase students' rating of faculty. These findings echo those noted by Marsh (1987) who suggested that higher workload levels and more difficult courses were positively associated with student ratings.

Of particular significance is that, by including student engagement as a moderator of student ratings of faculty, the focus, as noted by Coates (2005), is shifted back to students and their perceptions of their experience and their learning. Conversations about the quality of education come back to student classroom experiences and the extent to which students perceive they are engaged in their own learning. Given their role as participant observers in classrooms, students are in an excellent position to provide feedback regarding classroom teaching and overall performance of an instructor. They have a central stake in the quality of teaching and learning in the classroom. As Murray (1995) suggested, given the "symbiotic relationship between professors and students, it is not only in our best interests to respect what they can tell us about our teaching, but also in their best interests to assist us to improve our teaching" (p. 50).

The results of this research also suggest that those who are interested in student evaluation of their classroom experiences should consider constructing sound indicators of student engagement as part of the evaluation process, rather than spending time asking questions related to, for instance, whether or not the students liked the textbook. As shown by over fifty years of research on faculty evaluations and student ratings (e.g., Theall, Abrami, & Mets, 2001), students are eager to tell us what they think; we need to supply them with an appropriate, meaningful mechanism that includes information specific to the context of a course, such as student engagement.

As Abrami (2005) points out, promotion and tenure committees have a great responsibility for making life-altering decisions about their colleagues based on limited data regarding their performance in the classroom. Student evaluations are summative data and their use, especially across institutions but even within an institution, can have wide variability. He provides several suggestions for improving judgments about teacher effectiveness and several of these deal with examining the data more closely and in more disaggregated ways.

March (1987), recognizing the predominant use of student evaluations as summative data, noted that a central purpose guiding student evaluations of professors should, instead, be to provide feedback for the improvement of teaching. When the focus of teachers, and those who evaluate those teachers, is limited to only a part of the student rating instrument and how that one item may or may not be related to grades, the formative evaluative power of student feedback is lost. This is especially true when the relationship between grades and teacher ratings are strongly moderated by course contextual factors, such as the student's own engagement with the course material. Given the time and resources devoted to the collection of student ratings regarding the evaluation of teachers in higher education, imagine if student feedback and evaluating that feedback actually led to better teaching and enhanced student learning.

References

Abrami, P. L. (2005). Using teacher ratings forms to evaluate teaching: Doing a better job with what we've got. AERA symposium: *Valid faculty data: Are they any?* Annual Meeting of the American Educational Research Association, Montreal.

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.

Centra, J. A., & Gaubatz, N. B. (2000). Student perceptions of learning and instructional effectiveness in college courses. Research Report No. 9. *The Student Instructional Report II.* Princeton, NJ: Educational Testing Service.

Coates, H. (2005). The value of student engagement for higher education a quality assurance. *Quality in Higher Education, 11*(1), 25-36.

Eiszler, C. F. (2002). College students' evaluations of teaching and grade inflation. *Research in Higher Education, 43*(4), 483-501.

Ewell, P. T., & Jones, D. P. (1996). *Indicators of "good practice" in undergraduate education: A handbook for development and implementation.* Boulder, CO: National Center for Higher Education Management Systems.

Gillmore, G. M., & Greenwald, A. G. (1994). *The effects of course demands and grading leniency on student ratings of instruction.* Seattle, WA: University of Washington, Office of Educational Assessment.

Greenwald, A. G., & Gillmore, G. M. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology, 89,* 743-751.

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

Knapper, C. (2001). Broadening our approach to teaching evaluation. *New Directions For Teaching and Learning, 88,* 3-9.

Kuh, G. D., Kinzie, J. I., Schuh, J. H., Whitt, E. J., & Associates. (2005). *Student success in college: Creating conditions that matter*. San Francisco, CA: Jossey-Bass.

Marsh, H. W. (2001). Distinguishing between good (useful) and bad workloads on student evaluations of teaching. *American Educational Research Journal, 38,* 183-212.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11*, 253-288.

Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluation of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology, 92,* 202-208.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students. Vol. 2: A third decade of research.* San Francisco, CA: Jossey-Bass.

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.

Seldin, P. (Ed.). (1999). *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions.* Bolton, MA: Anker.

Shulman, L. S. (2002). Making differences: A table of learning. *Change, 34*(6), 36-45.

Spooren, P., Mortelmans, D., & Denekens, J. (2007). Student evaluations of teaching quality in higher education: Development of an instrument based on 10 Likert scales. *Assessment & Evaluation in Higher Education, 32*(6), 667-679.

Sproule, R. (2002). The underdetermination of instructor performance by data from the student evaluation of teaching. *Economics of Education Review, 21,* 287-294.

Theall, M., Abrami, P. A., & Mets, L. (Eds). (2001). The student ratings debate. Are they valid? How can we best use them? *New Directions for Institutional Research, No. 109.* San Francisco, CA: Jossey-Bass.

Theall, M., & Franklin, J. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction? *New Directions for Institutional Research, 27*(5), 45-56.

STEVEN CULVER is Associate Director of the Office of Academic Assessment at Virginia Tech. He is the author of several refereed articles and book chapters on gender equity, classroom evaluation practices, outcomes assessment, and student persistence in higher education. In addition, he has served as an evaluation consultant to such diverse organizations as the Education Ministry of Finland, the National Community College Center for Cooperative Education, the Junior Engineering Technical Society, the United States Department of Education, and the State Council for Higher Education in Virginia.