Student Ratings, Teacher Standards, and Critical Thinking Skills

Stephen L. Benton, Meixi Guo, Dan Li, and Amy Gross

The IDEA Center

Stephen L. Benton, The IDEA Center, and Emeritus Professor, Kansas State University, Manhattan, KS; Meixi Guo, Dan Li, and Amy Gross, The IDEA Center.

Correspondence concerning this paper should be addressed to Stephen L. Benton, Senior Research Officer, The IDEA Center, 301 South Fourth St., Suite 200, Manhattan, KS 6650

Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 2013.

Abstract

Student ratings data were analyzed from over 300 institutions using the IDEA Student Ratings System (SRS) during the years 2006 to 2010 (N = 297,180). Two interesting findings emerged from the study. First, students rate their progress on relevant learning objectives higher when they perceive the instructor has high achievement standards. Second, students rate the excellence of the instructor and course higher when they perceive that the instructor expects them to share responsibility for their own learning. These findings challenge the misconception that high course standards are associated with low student ratings.

Student Ratings, Teacher Standards, and Critical Thinking Skills

PURPOSE OF THE STUDY

Several authors have pointed out common misconceptions about student ratings of instruction that remain uncorroborated by research and make valid use of ratings difficult (Aleamoni, 1987; Benton & Cashin, 2012; Feldman, 2007; Kulik, 2001; Svinicki & McKeachie, 2011; Theall & Feldman, 2007). Among these are college students just want easy courses and ratings are just popularity contests. The implication is that instructors who want high student ratings should lower their standards. Two items on the IDEA Student Ratings of Instruction instrument assess student perceptions of the instructor's teaching standards: "The instructor expected students to take their share of responsibility for learning," and "The instructor had high achievement standards in this class." We examined the relationships between students' responses to these items and their ratings of progress on relevant learning objectives, their ratings of the excellence of the instructor and of the course.

We then examined teaching methods associated with student progress on critical thinking skills. Students reportedly do not make much progress in critical thinking as a result of the college experience (Arum & Roksa, 2011) in spite of the finding that faculty members agree almost unanimously that teaching students to think critically is the principal aim of undergraduate education (Bok, 2006; Higher Education Research Institute, 2009). We analyzed several possible explanatory variables that might be related to student progress on critical thinking skills: teacher standards and teaching methods.

THEORETICAL FRAMEWORK

Teacher Standards

As members of learning communities, students have certain privileges and responsibilities (Davis & Murrell, 1993); among these is the personal responsibility to learn, a belief held by many first-year college students (Devlin, 2002). Instructors can hold students responsible for learning by requiring them to complete assignments, to prepare for class, and to fulfill course requirements. When students are held accountable and achieve high performance standards, their self-efficacy increases and they are more motivated to pursue additional learning (Bandura, 1977). In fact, students' perceptions of teaching standards are positively correlated with their desire to take a course from the instructor (Hornbeak & Benton, 2010). More importantly, students whose instructors set high expectations tend to outperform other students (Marks, Doane, & Secada, 1998). In Study 1, we predicted that students would report more progress on relevant learning objectives if they perceived that the instructor set high achievement standards and expected students to share responsibility for learning. We also expected such teaching standards would be positively correlated with student ratings of the excellence of the instructor and the course.

Disciplinary Differences in Student Ratings

Feldman (1978) reviewed studies showing that courses in the humanities and arts receive higher ratings than those in social sciences, which in turn receive higher ratings than math and science. Others (Braskamp & Or y, 1994; Cashin, 1990; Centra, 1993, 2009; Hoyt & Lee, 2002b; Kember & Leung, 2011; Marsh & Dunkin, 1992; Sixbury & Cashin, 1995) found similar results. Although there is increasing evidence that ratings differ between disciplines, it is *not clear* why.

Cashin (1990) suggested some possible explanations. For example, some fields may be rated lower because they are more poorly taught; if so, then these differences do *not* require control. There is some evidence, for example, that mathematical/science courses tend to receive lower ratings (Centra, 2009; Hoyt & Lee, 2002b). Ratings are lower for the teaching styles of stimulating interest, fostering collaboration, and encouraging student involvement, which are related to overall measures of teaching excellence. Notably, students also rate the courses in these content areas more difficult than other courses, and they express less motivation to take the course (Hoyt & Lee, 2002b).

Another explanation for disciplinary differences is the sequential/hierarchical structure of content in some disciplines (Hativa, 2013b).. This is may especially be true of *hard* disciplines (see Biglan, 1973), meaning that they have a structure organized around a theory agreeable to all members of the field (e.g., engineering, chemistry). Because of such sequential structure, students must have a solid knowledge base in prior courses to succeed in subsequent ones. There may also be differences in the kinds of faculty and students attracted to certain disciplines. Some faculty in the hard sciences, for example, may be more attracted to a certain discipline because of research interests than for teaching opportunities. In addition, some students in specific fields might possess common attitudes toward and expectations about how courses should be taught.

Another possible explanation for disciplinary differences is the type of teaching methods employed in the classroom (Hativa, 2013b). Instructors in soft disciplines, for example, tend to exhibit a wider range of teaching behaviors than those in hard disciplines (Franklin & Theall, 1992). The authors found that instructors in the arts and humanities more frequently set objectives at the mid- and upper-levels of Bloom's taxonomy of cognitive objectives and used active teaching methods; whereas those in science, technology, engineering, and mathematics (STEM) courses rely more on lower-level objectives and employ lecture more predominately. Benton (2012) also reported that instructors in STEM fields employ teaching methods associated with student progress on relevant learning objectives less frequently than do those in non-STEM courses.

To understand whether the relationship between student perceptions of teacher standards and overall ratings depend on academic disciplines, we proposed the following questions.

RQ1: Does academic discipline moderate the relationship between student perceptions of teacher expectations/standards and progress on relevant course objectives?

RQ2: Does academic discipline moderate the relationship between student perceptions of teacher expectations/standards and overall ratings of teaching?

RQ3: Does academic discipline moderate the relationship between student perceptions of teacher expectations/standards and ratings of the course?

Critical Thinking

Critical thinking involves higher-order thinking skills such as evaluating, judging, problem solving, formulating inferences, and making reasoned decisions (Halpern, 1998). It is the kind of thinking that leads to desirable outcomes and is widely regarded as an important academic skill for college students (Bok, 2006; Higher Education Research Institute, 2009). Some assume that the development of critical thinking should emerge from the "college experience" (e.g., Arum & Roksa, 2011). Others argue that instruction and practice in critical thinking is required (Halpern, 1998). The IDEA Center's student learning model is based on the empirically supported premise that student progress on learning outcomes is positively related to the amount of emphasis the instructor places on those outcomes (Hoyt & Lee, 2002). Students report greater progress on objectives the instructor identifies as essential or important to the course (Benton, Webster, Gross, & Pallett, 2010). Students should not be expected to make progress on critical thinking, then, unless it is an important part of the course and unless the instructor does something to help students improve. Several specific teaching methods are highly correlated with student ratings of progress on the IDEA objective of "Learning to analyze and critically evaluate ideas, arguments, and points of view" (see Table 1). In Study 2, we employed stepwise regression to identify which ones are relatively most important.

RQ4: Which teaching methods are most strongly associated with student progress on critical thinking skills?

METHODS

Instrumentation

The IDEA Student Ratings of Instruction system is comprised of two forms: The *Faculty Information Form* and the *Student Rating Diagnostic Form*.

Faculty Information Form. Instructors complete the *Faculty Information Form* (FIF) for each course evaluated in the system. They first rate each of 12 learning objectives as 3 (*Essential*), 2 (*Important*), or 1 (*of Minor or No Importance*). Those identified as Essential or Important are considered "relevant objectives." Instructors then have the option of responding to contextual questions about the primary and secondary instructional approaches to the course; course requirements; whether any of several factors may have had a positive, negative, or neutral impact on students' learning; and the primary type of student enrolled in the class. Each campus determines the start and end dates for the survey completion. Instructors complete the FIF either online or on paper. The online version is delivered to faculty via e-mail.

Student Rating Diagnostic Form. The IDEA *Student Rating Diagnostic Form* is a 47-item instrument. Students indicate how frequently their instructor used each of 20 teaching methods, by responding 1 (*Hardly Ever*), 2 (*Occasionally*), 3 (*Sometimes*), 4 (*Frequently*), or 5 (*Almost Always*). The 20 teaching methods are organized into five teaching styles based on factor analysis (Hoyt & Lee, 2002). Using a 5-point scale (1 = no apparent progress, 2 = slight progress, 3 = moderate progress, 4 = substantial progress, 5 = exceptional progress), students also rate their progress on each of the same 12 learning objectives their instructor rated for importance.

Additional questions concern course characteristics, student characteristics, overall impressions of the course and instructor, other teaching methods and instructor standards. To measure students' self-reported progress, we calculated a weighted average of their ratings of the progress on objectives identified as "important" or "essential" (double weighted) by their instructors. To obtain a general evaluation of the course and instructor from students, we ask the extent to which the following two items reflect their attitudes on a 5-point scale (1 = *definitely false* and 5 = *definitely true*): "overall, I rate this instructor an excellent teacher" and "overall, I rate this course as excellent." Using the same Likert scale, two items tapped the instructor standards: "The instructor expected students to take their share of responsibility for learning" and "The instructor had high achievement standards in this class."

Samples of the faculty and student forms may be found at http://www.theideacenter.org/services/student-ratings/sample-forms-student-ratings-instruction. Students may complete the survey either online or on paper. Four survey delivery methods are available online: survey links available through a Blackboard® Building Block, e-mail, the course website, or a combination of all three. Students are restricted to one submission.

Coding Disciplines as Hard-Soft and Pure-Applied

Respondents can indicate their discipline by filling out a four-digit number representing their department. A list of 327 discipline/sub-disciplines can be found at http://theideacenter.org/DisciplineCodes. One author created two dichotomous variables respectively for the hard-soft and pure-applied characteristics and rated each discipline as either hard or soft and either pure or applied. Discipline that have been covered in existing literature (Ballantyne et al., 1999; Biglan, 1973b; 1973a; Neumann et al., 2002) were coded consistently with Biglan’s (1973a; 1973b) scheme and others were rated by the research’s judgment. A small number of disciplines were not coded due to their interdisciplinary characteristics. After the initial coding was completed, another author reviewed the result and discussed with the other code when disagreements arose. Necessary changes to the coding were made after the discussion. Both researchers agreed on classification of 95.4% of the disciplines, indicating high inter-rater reliability. Table 2 displays the frequency of classes by course format and the hard-soft, pure-applied dimensions.

DATA SOURCE

Student ratings data were analyzed from over 300 institutions using the IDEA Student Ratings System (SRS) during the years 2006 to 2010 (N = 297,180). Prior to conducting analyses, we first removed classes with fewer than 10 responses and instructors identified as first-time users of IDEA. In addition, institutional frequencies were checked, and the condition that all institutions each contributed no more than approximately 5% of classes was satisfied.

RESULTS

Because of the exceptionally large number of classes in the dataset, statistically significant results do not necessarily indicate practical significance. Therefore, R^2, a measure of explained variance, is reported throughout as an estimator of effect size rather than reporting tests of significance.

Student Progress on Relevant Objectives

The first research question concerned whether disciplinary categories moderated the relationship between student perceptions of teacher expectations/standards and student self-reported progress on relevant objectives. Table 2 presents means, standard deviations, and correlations for all variables in the models. For the moderated regression analysis, on the first step we tested main effects for discipline structure (hard, soft), application (pure, applied), teacher expectations, and achievement standards. This model was statistically significant, F (4,471220) = 103,843, $p < .0001$, $R^2 = .469$. Standardized beta coefficients showed that having high achievement standards (β = .45, $p < .0001$) was relatively more important than expecting students to share in the responsibility for learning (β =. 26, $p < .0001$). Discipline structure (β =. -.042, $p < .0001$) and application (β =. -.038, $p < .0001$) were relatively less important predictors.

On the second step, we added to the initial model the following two-way interactions: Structure x Teacher Expectations, Structure x Teacher Standards, Application x Teacher Expectations, and Application x Teacher Standards. However, collinearity statistics were exceedingly high. We then attempted to test only one disciplinary interaction term at a time, but collinearity statistics remained high. Even so, the model containing the moderators only added less than 1% of the variance.

Overall Ratings of the Instructor

We then examined whether disciplinary categories moderated the relationship between student perceptions of teacher expectations/standards and overall ratings of the instructor. Using the same models described above, the main effects were significant on the first step, F (4,471312) = 79927.41, $p < .0001$, $R^2 = .404$. Expecting students to share in their responsibility for learning ($\beta = .35$, $p < .0001$) and having high achievement standards (β =. 32, $p < .0001$) contributed more to the model than either discipline structure ($\beta = -.06$, $p < .0001$) or application ($\beta = .05$, $p < .0001$). On the second step, we again encountered collinearity problems, and the model that included moderators explained less than 1% of additional variance in instructor ratings.

Overall Ratings of the Course

In Step 1, the main effects explained significant variance in overall ratings of the course, F (4,471312) = 88518.69, $p < .0001$, $R^2 = .429$. Teacher expectations (β =. 36, $p < .0001$) and achievement standards (β =. 31, $p < .0001$) were again most important for explaining variance in course ratings. Discipline structure (β = - .08, $p < .0001$) and application (β = -. 06, $p < .0001$) were relatively unimportant. Multicollinearity was again extreme in the second step, which explained less than 1% additional variance.

Taken together, the analyses in Study 1 revealed that students' perceptions of teacher characteristics are related to the progress they report and their overall impressions of the teaching and course. Students who perceive the instructor expects them to share in responsibility for learning and sets high achievement standards are more likely to make progress in the course and assign high ratings.

Critical Thinking and Teaching Methods

We conducted a step-wise regression analysis to identify teaching methods that significantly predict students' progress on critical thinking. The dependent variable for this analysis was student self-ratings of progress on the objective "Learning to analyze and critically evaluate ideas, arguments, and points of view." As expected, students reported more progress when instructors rated this objective as essential or important for the course, ($R^2 = .05$, $p < .0001$). The explanatory variables entered into a stepwise regression were those listed in Table 1. The teaching methods in Table 1 have consistently been the ones most highly correlated with progress on this objective (Hoyt & Lee, 2002). The teaching method of "asking students to share ideas and experiences with others whose backgrounds are different from their own" entered on the first step ($R^2 = .61$, $p < .0001$). This was followed on the second step by the method of "stimulating students to intellectual effort beyond that required by most courses" (partial $R^2 = .11$, $p < .0001$). On the third step, "giving projects, tests or assignments that required original or creative thinking" explained an additional 1% of the variance ($R^2 = .01$, $p < .0001$). No other variables had a partial R^2 that approached a magnitude of at least .01 and consequently were excluded from the model. Table 3 presents means, standard deviations, and correlations for all variables retained in the model. With respect to critical thinking, then, students reported the most progress when the instructor asked students to share ideas with others and stimulated them to intellectual effort.

SIGNIFICANCE OF THE STUDY

Several interesting findings emerged from this study. First, students rate their progress on relevant learning objectives higher when they perceive the instructor has high achievement standards. Second, students rate the instructor and course excellence higher when they perceive that the instructor expects them to share responsibility for their own learning. These findings challenge the unsupported claim that "If [faculty] raise course demands on their students but their peers do not, they will potentially be disadvantaged by course evaluations in which students express dissatisfaction" (Arum & Roksa, p. 134). Previous research established that college students have a stronger desire to take a course when the instructor sets high achievement standards (Hornbeak & Benton, 2010). Moreover, students whose instructor has high expectations for achievement tend to outperform other students (Marks et al., 1998). So, the current findings challenge the common misconception that instructors who have easy courses get higher ratings.

Another finding is that two teaching methods emerged as most important for helping students to make progress on critical thinking skills. First, instructors who ask students to share ideas and experiences with others whose backgrounds are different from their own have students who report greater progress on critical thinking. This is consistent with Richard Paul's (1987) advocacy of *dialogical thinking*, in which students consider problems and solutions from different points of view, as a means for teaching critical thinking. One way to do this is to create collaborative learning activities in which students feel safe expressing ideas and beliefs without fear of ridicule (King, 2005).

Second, teachers should stimulate students to intellectual effort beyond that required by most courses. They can, for example, challenge students just beyond their current capability and require them to engage in real-world applications (McClure, 2005).

Taken together, the current findings can be summarized as follows: a) students give higher overall ratings of their progress, the instructor, and course when the teacher expects more of them; and b) students report greater progress on critical thinking skills when the instructor asks them to share experiences with diverse peers and challenges them to greater intellectual efforts.

References

Aleamoni, L. M. (1987). Typical faculty concerns about student evaluation of teaching. In L. M. Aleamoni (Ed.), *Techniques for evaluating and improving instruction: New Directions for Teaching and Learning,* No. 31 (pp. 25-31). San Francisco: Jossey-Bass.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses.* Chicago, IL: University of Chicago Press.

Bandura, A. (1977). *Social learning theory.* Englewood Cliffs, NJ: Prentice Hall.

Benton, S. L., & Cashin, W. E. (2012). *IDEA Paper No. 50: Student ratings of teaching: A summary of research and literature*. Manhattan, KS: The IDEA Center.

Benton, S. L., Webster, R., Gross, A., & Pallett, W. H. (2010). *IDEA Technical Report No. 16: An analysis of IDEA Student Ratings of Instruction using paper versus online survey methods, 2002-2008 data.* Manhattan, KS: The IDEA Center.

Bok, D. (2006). *Higher learning.* Cambridge, MA: Harvard University Press.

Davis, T.M., & Murrell, P.H. (1993). A structural model of perceived academic, personal, and vocational gains related to college student responsibility. *Research in Higher Education*,Vol. 34, No. 3, 267-289.

Devlin, M. (2002). Taking responsibility for learning isn't everything: A case for developing tertiary students' conceptions of learning. *Teaching in Higher Education*, Vol. 7, No. 2, 125-138.

Feldman, K. A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry & J. C. Smart (Eds.), *The Scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 93-129). Dordrecht, The Netherlands: Springer.

Halpern, D. (1998).

Higher Education Research Institute. (2009). *The American College Teacher: National Norms for 2007-2008*. Los Angeles: HERI, University of California Los Angeles.

Hornbeak, J. L., & Benton, S. L. (2010, May). *Teaching methods and course characteristics related to college students' desire to take a course.* Paper presented at the annual meeting of the American Educational Research Association, Denver.

Hoyt, D. P., & Lee, E. (2002). *Technical Report No. 12: Basic data for the revised IDEA system.* Manhattan, KS: The IDEA Center.

King, J. (2005). IDEA Item #16: "Asked students to share ideas and experiences with others whose backgrounds and viewpoints differ from their own." POD-IDEA Center Notes. Manhattan, KS: The IDEA Center.

Kulik, J. A. (2001). Student ratings: Validity, utility, and controversy. In M. Theall, P. C. Abrami, & L. A. Mets (Eds.). The s*tudent ratings debate: Are they valid? How can we best use them? New Directions for Institutional Research*, No. 109

(pp. 9-25). San Francisco: Jossey-Bass.

Marks, H., Doane, K., & Secada, W. (1998). Support for student achievement. In F. Newmann et al. (Eds.), *Restructuring for student achievement: The impact of structure and culture in 24 schools.* San Francisco: Jossey-Bass.

McClure, N. C. (2005). IDEA Item #8: "Stimulated students to intellectual effort beyond that required by most courses." POD-IDEA Center Notes. Manhattan, KS: The IDEA Center.

Paul, Richard W. (1987). In *Teaching thinking skills: Theory and practice*, by Paul, Richard W., 127-148. New York, NY, US: W H Freeman/Times Books/ Henry Holt & Co, 1987.

Svinicki, M. & McKeachie, W. J. (2011). *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers*. (13th ed.). Belmont, CA: Wadsworth.

Theall, M., & Feldman, K. A. (2007). Commentary and update on Feldman's (1997) "Identifying exemplary teachers and teaching: Evidence from student ratings." In R. P. Perry and J. C. Smart (Eds.), *The teaching and learning in higher education: An evidence-based perspective (pp. 130-143)*. Dordrecht, The Netherlands: Springer.

Table 1

Teaching Methods Associated with Progress on Critical Thinking

TM2 Found ways to help student answer their own questions

TM7 Explained the reasons for criticisms of students' academic performance

TM8 Stimulated students to intellectual effort beyond that required by most courses

TM13 Introduced stimulating ides about the subject

TM15 Inspired students to set and achieve goals which really challenged them

TM16 Asked student to share ideas and experiences with others whose backgrounds and viewpoints differ from their own

TM19 Gave projects, tests, or assignments that required original or creative thinking

Note. All methods correlated $r = .60$ or above with the objective "Learning to analyze and critically evaluate ideas, arguments, and points of view"

Table 2

Frequency of Courses by Hard-Soft Dichotomy and Pure-Applied Dichotomy

	Application	
Structure	Applied	Pure
Soft	184,061	159,106
Hard	57,200	70,988

Note. N = 490,333. 18,978 classes were excluded from this table because they could not be classified as hard-soft or pure-applied.

Table 3

Means, Standard Deviations, and Correlations for all Explanatory and Dependent Variables in Models 1 and 2

Variable	1	2	3	4	5	M	SD
1. PRO		.79	.81	.62	.66	4.10	.45
2. Teacher			.86	.61	.60	4.27	.60
3. Course				.62	.61	4.04	.58
4. Responsibility					.81	4.34	.32
5. Standards						4.19	.39

Note. All correlations significant at $p < .0001$. PRO = progress on relevant objectives; Teacher = excellence of teacher; Course = excellence of course; Responsibility = expecting students to take their share of responsibility for learning; Standards = having high achievement standards in the class.

Table 4

Means, Standard Deviations, and Correlations for all Explanatory and Dependent Variables in Model 3

Variable	1	2	3	4	M	SD
1. Critical thinking		.77	.76	.76	3.83	.58
2. TM16			.66	.78	3.91	.72
3. TM8				.72	4.08	.53
4. TM19					4.09	.59

Note. All correlations significant at $p < .0001$. TM16 = asking students to share ideas and experiences with others whose backgrounds and viewpoints differ from their own; TM8 = stimulating students to intellectual effort beyond that required by most courses; TM19 = giving projects, tests, or assignments that required original or creative thinking.