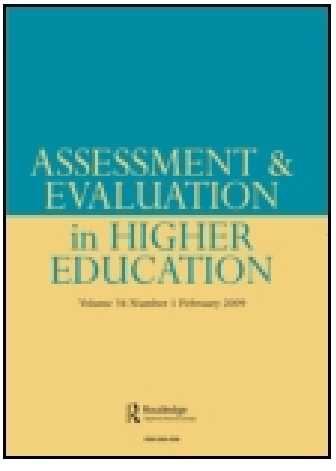

Assessment & Evaluation in Higher Education

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/caeh20

Validity of student self-reported ratings of learning

Stephen L. Benton [a] , Dennis Duchon [b] & William H. Pallett [a]

[a] The IDEA Center , Manhattan , KS , USA

[b] Department of Management , University of Nebraska-Lincoln , Lincoln , NE , USA

Published online: 16 Nov 2011.

To cite this article: Stephen L. Benton , Dennis Duchon & William H. Pallett (2013) Validity of student self-reported ratings of learning, Assessment & Evaluation in Higher Education, 38:4, 377-388, DOI: 10.1080/02602938.2011.636799

To link to this article: http://dx.doi.org/10.1080/02602938.2011.636799

Assessment & Evaluation in Higher Education, 2013
Vol. 38, No. 4, 377–388, http://dx.doi.org/10.1080/02602938.2011.636799

Validity of student self-reported ratings of learning

Stephen L. Benton[a]*, Dennis Duchon[b] and William H. Pallett[a]

[a]*The IDEA Center, Manhattan, KS, USA;* [b]*Department of Management, University of Nebraska-Lincoln, Lincoln, NE, USA*

This study examined the relationship between individual student self-reported ratings of progress on relevant learning objectives and performance on exams administered during a college course. Across three sections of the same course taught by a single instructor, 188 students rated themselves at the end of the course on two objectives identified by the instructor as either essential or important. They also rated themselves on 10 other objectives the instructor identified as having minor or no importance. Self-ratings on course-relevant objectives correlated significantly and positively with four out of five exams and the course total, whereas ratings on irrelevant objectives did not. Students who rated their progress as either exceptional or substantial generally performed better on course examinations than those who rated their progress as moderate or less. These findings support the validity of student self-reported ratings of learning.

Keywords: student ratings; learning objectives; college students; self-report

Student ratings of instruction, which have been used for decades to evaluate college teaching, have long been the focus of intense debate (e.g. Remmers 1930). The key question has been to what extent student ratings measure some aspect of teaching effectiveness. The stumbling block is how one defines effective teaching. Some have made the case that one measure of teaching effectiveness is how much students have learned in a course (e.g. Cashin 1995; Cohen 1981; Feldman 1989; Marks, Fairris, and Beleche in press; McKeachie 1979). Other things being equal, effective teaching – no matter where it occurs should lead to greater student learning (Cashin 1995). However, students typically complete ratings anonymously, which makes it difficult to correlate individual student learning with evaluations of the course. Studies that link individual student ratings with learning outcomes are, therefore, needed (Marks, Fairris, and Beleche in press). The current study does this by examining the relationship between individual student self-reported ratings of progress on course-relevant learning objectives and performance on exams administered during a college course.

The relationship between student ratings and learning has been addressed in studies that correlate some measure of student achievement (e.g. exam performance) with course evaluations across multiple sections of the same course. Typically, different instructors teach multiple sections, using the same syllabus, textbook and external final exam. In their reviews of such studies, Cohen (1981) and Feldman (1989) reported average correlations (Pearson's *r*) exceeding .40 between an

*Corresponding author. Email: bill@theideacenter.org

external final exam score and student ratings of self-reported learning, overall evaluation of the course, overall evaluation of the instructor, teacher skill, course preparation, teacher structure and instructor clarity. The classes that received the highest ratings tended to be ones in which students reported learning the most.

More recently, Marks Fairris, and Beleche (in press) correlated disaggregated student ratings with performance on a common external final exam across multiple sections of a remedial college course. Sections were taught by different instructors, but the text, number of assignments and point system for grading were the same. All students took a common post-test that was centrally graded. The authors focused specifically on student responses to the question, 'The course overall as a learning experience was excellent'. Marks et al. found a consistently positive relationship between ratings on this item and scores on the common final exam, controlling for student demographic characteristics, ability and pre-test scores. So, again, course ratings correlated positively with measures designed to assess student learning.

Validity and reliability of student self-reports of learning

The current study focuses exclusively on one measure of teaching effectiveness: student self-ratings of progress on objectives the instructor identifies as essential or important. The notion that instruction can be evaluated in terms of student progress on relevant course objectives has a long history (e.g. McKeachie 1979). Such a measure has value because it reflects the students' judgements of how much they believe they have learned about what the instructor identifies as important knowledge for any given course.

Learning quality varies from one course to the next. Some teachers may emphasise higher-order objectives (e.g. analysing and critically evaluating ideas) whereas others may find lower-order objectives (e.g. gaining factual knowledge) more appropriate. Regardless of which type of learning is emphasised, students tend to report more progress on objectives the instructor identifies as important or essential than of minor or no importance (Hoyt and Perera 2000). The validity of student self-assessments of learning, however, depends on the degree to which instructors accurately identify relevant learning objectives and students make accurate estimates of their learning. When instructors articulate specific, desirable student learning outcomes, it makes the process more feasible.

Although there is a significant correlation between faculty ratings of the importance and student ratings of learning on the same objective, past studies (Hoyt 1973; Hoyt and Lee 2002) relied upon the average class ratings, not the ratings of individual students. The current study investigated the relationship between individual student self-reported learning and achievement as indicated by scores on objectively scored examinations. We hypothesised that exam performance would be positively correlated with student self-reported progress on learning objectives identified by the instructor as relevant to the course but unrelated to objectives identified as less relevant. This follows from McKeachie's (1979, 385) definition of teaching effectiveness as 'the degree to which one has facilitated student achievement of educational goals'.

We undertook several measures to enhance the credibility of the study. Students' identities were protected so that the ratings would not affect their grade. To reduce the threat of researcher expectancy (Rosenthal 1969), a neutral party (the faculty

member's department chair) administered, collected, and mailed the ratings for processing. The instructor and researchers were, therefore, blind to individual and aggregate ratings until all exams were scored and grades were submitted. To ensure objectivity in scoring, all exams were multiple-choice format, and all but the final exam were completed before students rated the course and instructor. As a control against reactivity, in the form of 'hypothesis guessing' (Krathwohl 1998), students rated their progress on objectives the instructor deemed relevant *and* less relevant to the course. All objectives were weighted for importance based on ratings the instructor completed prior to teaching the class.

Purpose and predictions

The purpose of this study was to examine the relationship between individual student self-reported ratings of progress on relevant learning objectives, as determined by the instructor, and performance on exams administered during the course. The design of the study enabled the researchers to see whether students who achieve more assign higher progress ratings, regardless of the course section. We proposed the following hypotheses. First, performance on course examinations should be significantly and positively correlated with student ratings of progress on relevant learning objectives. The instructor in this study identified two objectives as relevant to the course that were highly inter-correlated. Therefore, we expected that student ratings of progress on each would be positively correlated with exam performance on items tied to both objectives. Second, exam performance should not be significantly correlated with ratings of progress on objectives the instructor identified as having little or no importance to the course (i.e. less relevant objectives). Third, students who rate their progress as either exceptional or substantial on relevant objectives should score higher on relevant exam content than those rating their progress as moderate (or lower), controlling for prior achievement. Fourth, there should be no differences in exam performance between students who rated their progress as exceptional, substantial or moderate on less relevant objectives.

Method

Participants

Faculty from a department of management in the south-western US and staff from The IDEA Center collaborated on the study. Participating students were in three sections of an undergraduate management course, all taught by the same instructor. A total of 188 students completed all five examinations and the course evaluation.

Instrumentation

Faculty Information Form

The IDEA Student Ratings System (http://www.theideacenter.org) was developed through a grant from the Kellogg Foundation in 1975. The IDEA Center is a non-profit organisation that has as part of its mission supporting the improvement of learning and teaching through the use of its diagnostic student ratings instrument.

The IDEA system relies principally on the assumption that instructional effectiveness is best indicated by students' self-reported progress on learning objectives the instructor identifies as essential or important. On the Faculty Information Form

(FIF), instructors rate each of 12 learning objectives as either 'Essential', 'Important', or of 'Minor or No Importance'. The learning objectives were initially developed by reviewing two taxonomies of educational objectives (Bloom 1956; Krathwohl, Bloom, and Masia 1964) and then reducing the number to a smaller set of objectives endorsed by professors who had won outstanding teaching awards (Hoyt and Cashin 1977). Subsequently, three of the original 10 were dropped and five new ones were added in keeping with changes in higher education (Hoyt et al. 1999). The current 12 objectives appear in the appendix.

Student Rating Form

Students indicate their progress on the same 12 learning objectives the instructor rated for importance on the FIF. They use the scale of 1 (*No apparent progress*); 2 (*Slight progress; I made small gains on this objective*); 3 (*Moderate progress; I made some gains on this objective*); 4 (*Substantial progress – I made large gains on this objective*), and 5 (*Exceptional progress; I made outstanding gains on this objective*). Students also indicate how frequently their instructor used each of 20 teaching methods, using a scale of 1 (*Hardly Ever*), 2 (*Occasionally*), 3 (*Sometimes*), 4 (*Frequently*) and 5 (*Almost Always*). Additional items concern course characteristics, student characteristics (e.g. work habits and motivation), and overall ratings of the course and instructor.

Hoyt and Lee (2002) computed split-half reliabilities for each item on the Student Rating Form. Classes that contained 13–17 respondents were randomly divided into two groups, and the means were computed for each group and correlated. The Spearman–Brown Prophecy Formula was then used to estimate the reliabilities for class size averages of 12.5, 24.5, 42.5 and 60 (which corresponded to class sizes of 10–14, 15–34, 35–49 and 50+). Average reliability coefficients for individual items ranged from .78 for class sizes of 10–14 students to .94 for enrollments of 50 or more. Standard errors of measurement ranged from .27 to .14 on the five-point scale. These statistics provide evidence to support the degree of reliability in individual items.

An indirect test of the validity of the IDEA ratings involves correlating students' reported progress on each objective with the instructors' ratings of the importance of those objectives. The highest correlations should be found in ratings of the same objectives (Hoyt 1973). Indeed, Hoyt (1973) found that the average correlation ($r = .32$, $p < .0001$) between ratings of the same objectives was higher than that of different objectives ($r = .02$, $p < .30$).

Additional validity evidence is found in correlations between students' self-ratings of progress on the learning objectives and their ratings of how frequently the instructor emphasised 20 specific teaching methods. For example, Hoyt and colleagues (Hoyt and Lee 2002; Hoyt and Perera 2000) found that the teaching method most highly correlated with progress on the learning objective of 'Team skills' was 'Formed teams or discussion groups to facilitate learning' ($r = .86$, $p < .0001$). Similarly, progress on 'Learning to find and use resources for answering questions or solving problems' was most highly correlated with the method of 'Encouraged students to use multiple resources to improve understanding' ($r = .82$, $p < .0001$). How frequently an instructor emphasises a specific teaching method routinely correlates positively with students' self-reported progress on an objective conceptually related to that method. Because of this, in its individual reports to instructors, The IDEA

Center makes recommendations about which methods are most highly correlated with student ratings on relevant course objectives. Instructors are encouraged to either retain current use of a teaching method that is highly correlated with a specific learning objective (if they use the method frequently) or increase its use (if they use the method infrequently).

Course examinations

Students completed five multiple-choice course examinations. Some exam items were drawn from the textbook's supplemental test bank, whereas others were created by the instructor. All items were tied to course content related to two relevant IDEA objectives: 'Gaining factual knowledge (terminology, classifications, methods, trends)' and 'Learning fundamental principles generalisations or theories'. The instructor in collaboration with the department chair classified each item on these examinations as 'Factual Knowledge' or 'Principles and Theories'. The number of points possible varied across tests. The correlations reported in the results section address the exams' validity and reliability.

Procedures

Of the 12 learning objectives listed on the IDEA form, the instructor identified two as being relevant for all three sections of the course: 'Gaining factual knowledge (terminology, classifications, methods, trends)' and 'Learning fundamental principles, generalisations or theories'. The instructor used the same syllabus, textbook and lecture materials for all three sections. At the end of the semester students rated their learning on all 12 objectives. The instructor scored all exams before students completed the ratings, and he had no knowledge of the results of the ratings until after grades were submitted.

The chair of the management department administered the IDEA Student Ratings of Instruction form at the end of the semester. He described the proposed study and informed students that their responses could not be anonymous because they were to be matched with other institutional and class data. He nonetheless assured students their identity would be protected so that the ratings could not influence their grades. Informed consent was obtained, and nearly everyone who attended class the day the IDEA survey was administered participated in the course evaluation.

Results

Stability of exam scores

To assess the reliability (i.e. stability) of exam scores, we correlated student summative scores on factual knowledge items with principles and theories items *within* each exam, which resulted in the following significant correlations ($p < .0001$): Exam 1 ($r = .49$), Exam 2 ($r = .72$), Exam 3 ($r = .60$), Exam 4 ($r = .60$) and Exam 5 ($r = .61$). Stability was also found *between* exam scores on 'Factual Knowledge' and 'Principles and Theories' items. Table 1 shows that the degree of test-to-test consistency in performance on each objective was no higher (and was sometimes lower) than the correlation between scores on the two components within each testing occasion (reported above). The mean correlation between corresponding exam

Table 1. Pearson correlations between student examination scores and ratings of progress on IDEA objectives.

	IDEA student ratings of progress		
Exam #/emphasis	Factual knowledge	Principles/ theories	Average *r* other 10 objectives
#1-Factual knowledge	.30[a]	.34[a]	.09
#1-Principles theories	.23[a]	.27[a]	.03
#1-Total	.31[a]	.37[a]	.08
#2-Factual knowledge	.11	.05	−.11
#2-Principles theories	.09	.13	−.11
#2-Total	.11	.08	−.11
#3-Factual knowledge	.28[a]	.29[a]	.07
#3-Principles theories	.27[a]	.26[a]	.04
#3-Total	.31[a]	.30[a]	.06
#4-Factual knowledge	.19[b]	.24[a]	−.07
#4-Principles theories	.19[b]	.24[a]	−.07
#4-Total	.21[a]	.27[a]	−.07
#5-Factual knowledge	.40[a]	.31[a]	−.01
#5-Principles theories	.33[a]	.29[a]	.09
#5-Total	.41[a]	.34[a]	.04
Course factual knowledge	.32[a]	.30[a]	−.02
Course principles and theories	.30[a]	.33[a]	−.03
Course total	.32[a]	.32[a]	−.07

Notes: [a]Correlation is significant at the .01 level (Two-tailed); [b]Correlation is significant at the .05 level (Two-tailed).

components was $r = .49$ ($p < .0001$) for factual knowledge and $r = .38$ ($p < .0001$) for principles and theories. Student performance within and across examinations was, therefore, moderately stable. This again confirms the redundancy in the two learning objectives and the exam items, which was expected, given students' inability to distinguish their performance on these two objectives and the similarity in the exam content.

The central question of the study was whether students' ratings of their learning would correlate positively with objective (test-based) assessments of achievement. The most straightforward way to examine this question was to correlate self-ratings with examination scores. Of the 188 students, only seven rated 'Factual Knowledge' and only eight rated 'Principles and Theories' as '1' (*No apparent progress*) or '2' (*Slight progress*). For the purposes of this study, students who rated their progress on an objective as '1', '2' or '3' (*Moderate progress*) were placed in the same group. Table 2 displays in the first two columns the correlations between students' self-reported ratings of learning on the two 'relevant' objectives and each of three scores from the five classroom examinations ('Factual knowledge' score, 'Principles and theories' score, and a 'Total' score, the sum of the previous two). For comparison purposes, the average correlations of student ratings on the 10 'irrelevant' IDEA objectives (objectives not emphasised in this course) are shown in the third column.

Exams 1, 3, 4 and 5 scores were significantly related to self-reported learning on the two objectives the instructor identified as essential or important (see Table 2). As hypothesised, student ratings of progress on factual knowledge were positively correlated with their performance on factual knowledge questions

Table 2. Intercorrelations between factual knowledge and principles and theories subtests across five exams.

Exams	1	2	3	4	5
1		.45	.57	.60	.48
2	.32		.45	.42	.46
3	.41	.41		.57	.48
4	.34	.34	.46		.40
5	.36	.35	.46	.32	

Notes: Factual knowledge correlations are above diagonal; principles and theories are below diagonal. All correlations are significant at the .01 level.

for the entire course ($r = .32$, $p < .0001$). The same was true for principles and theories ($r = .33$, $p < .0001$). This compares to average correlations of $r = -.02$ ($p < .39$) and $-.03$ ($p < .34$), respectively, for all other objectives. Furthermore, for the 10 objectives that were not relevant to this class (i.e. not chosen by the instructor), none of the correlations between examination scores and student ratings of learning was significant. Based on these results, Hypotheses 1 and 2 were confirmed.

The correlations in Table 2 lend support to the argument that student self-reported ratings of learning are a useful proxy for performance-based measures of learning. Although the correlations between exam scores and relevant objectives were far from a perfect 1.0, they represent a small to medium effect (Cohen 1977) and were impressive when one considers the limited range on the student ratings scale. When coupled with the correlations between exam scores and irrelevant objectives, these results support the convergent and discriminant validity of the classroom exams.

Hypothesis 3 posited that students who rated their progress as substantial or exceptional would score higher on relevant exam content than those rating their progress as moderate or lower. Prior to testing this hypothesis, we conducted ANOVAs between course sections on nine individual student achievement and aptitude measures (SAT Verbal, SAT Math, SAT Composite, ACT English, ACT Math, ACT Reading, ACT Science, ACT Total and cumulative institutional GPA). To control for Type I error when conducting multiple tests, we made Bonferroni adjustments ($\alpha = .05/9 = .006$). No significant differences were found across sections. We also compared course sections on the dependent variables of interest (i.e. exam scores) and found no statistically significant differences across sections. We, therefore, collapsed across sections for all analyses. Cumulative GPA was used to estimate prior learning, largely because it was available for most subjects and was free from the problem of institutional differences in grading practices that a measure like 'transfer GPA' would introduce.

Two multivariate analyses of covariance (MANCOVA) were conducted to determine if 'adjusted' exam scores (those that took into account previous grades at the institution) varied for students who made different ratings of their learning. The dependent measures were total scores on factual knowledge (Total Objective 1) and principles and theories (Total Objective 2) items. For the first MANCOVA, the independent variable was student self-rating on factual knowledge (exceptional, substantial and moderate). The multivariate test was significant for self-rating on factual

knowledge, *Wilks's* $\lambda = .91$, $p < .001$, $\eta^2 = .05$. Cumulative GPA, the covariate, explained 25% ($\eta^2 = .25$) of the variance in the dependent variables. Univariate tests were significant for Total Objective 1, F (2,133) = 5.72, MSE = 654.67, $p < .004$, $\eta^2 = .08$; and Total Objective 2, F (2,133) = 6.21, MSE = 206.51, $p < .003$, $\eta^2 = .09$. Table 3 presents adjusted means and standard deviations, along with results of Bonferroni post hoc tests. Students who rated their progress on factual knowledge as either substantial or exceptional outperformed those who rated it as moderate on both Total Objectives 1 and 2.

A second MANCOVA was conducted using student self-ratings on principles and theories (Objective 2) as the grouping variable. The multivariate test was significant, *Wilks's* $\lambda = .93$, $p < .041$, $\eta^2 = .04$. Univariate tests were significant for Total Objective 1, F (2,133) = 3.62, MSE = 674.25, $p < .029$, $\eta^2 = .05$; and Total Objective 2, F (2,133) = 5.08, MSE = 209.77, $p < .007$, $\eta^2 = .07$. Students who rated their progress on principles and theories as exceptional outperformed those who rated it as moderate on both Total Objectives 1 and 2 (see Table 3).

In each of the above analyses the average exam score increased as students' self-ratings of learning increased. As shown in Table 3, across both factual knowledge and principles and theories, students in the moderate group (i.e. reporting either moderate or slight or no apparent progress) had descriptively lower exam scores than those reporting higher levels of learning. The difference between students rating their progress as substantial vs. exceptional was not statistically significant, but the average for those rating their progress as exceptional consistently exceeded that of substantial, and those rating progress as substantial exceeded that of moderate. In general, then, Hypothesis 3 was confirmed.

Hypothesis 4 stated that there would be no differences in exam performance between students who rated their progress as exceptional, substantial or moderate on less relevant objectives. To test this hypothesis, we applied MANCOVA on the 10 learning objectives identified by the instructor as having 'no more than minor importance'. No tests reached the $\alpha = .05$ level of significance. Student ratings of progress on objectives less relevant to the course were, therefore, unrelated to exam performance.

Table 3. Adjusted means and standard deviations for total subtest scores by student self-ratings of progress on factual knowledge and principles and theories.

	Total scores				
	Factual knowledge			Principles and theories	
Self-rating group	*n*	*M* (*adj*.)	*SD*	*M* (*adj*.)	*SD*
Factual knowledge					
Moderate	29	253.01_a	28.12	109.54_a	17.15
Substantial	46	269.62_b	33.46	119.45_b	16.09
Exceptional	62	271.91_b	26.18	120.53_b	15.72
Principles and theories					
Moderate	28	256.09_a	25.43	110.68_a	13.64
Substantial	61	267.88_{ab}	31.93	118.12_{ab}	17.12
Exceptional	48	272.65_b	27.89	121.67_b	16.09

Note: Means with different subscripts are significantly different at the .05 level.

Conclusion

This study provides support for the validity of student self-reported ratings of learning. Average exam scores were significantly and positively correlated with student ratings of learning on objectives identified as either essential or important by the faculty member (factual knowledge and principles and theories). Exam scores did not correlate significantly with student ratings of progress on less relevant objectives. When examination results were adjusted on the basis of past academic performance, students who rated their progress as exceptional or substantial generally outperformed those who rated it as moderate or lower. No such differences were found between students who varied in their ratings of progress on less relevant objectives.

These results support previous findings that students' ratings of their own learning correlate positively with the instructor's measure of how much they have actually learned (Cohen 1981; Feldman 1989; Marks Fairris, and Beleche in press). They also indicate that students are capable of assessing what they have learned at the time they complete course ratings. Although the current study involved only a single course across multiple sections, it supported the argument that an important measure of teaching effectiveness is how much students believe they have learned in a course (e.g. Cashin 1995; Cohen 1981; Feldman 1989; Marks Fairris, and Beleche in press; McKeachie 1979).

The current findings provide evidence for the validity of student ratings based on assessment–criterion relationships (Messick 1995; Ory and Ryan 2001). We found a relationship – in the expected direction – between student ratings and a criterion external to the assessment. This supports Marsh's (1984) 'validity hypothesis' that students rate on the basis of how much they believe they have learned. As McKeachie (1979) pointed out, significant correlations between student ratings and achievement measures (i.e. exam performance) are not necessarily a sign of grading leniency or invalidity. On the contrary, one should expect such relationships because students who report greater learning should also perform better in the class.

We acknowledge several limitations of the current study. First, the data were gathered from only one course taught by a single instructor, which limits generalisability. Additional studies should be conducted across multiple courses, disciplines and institutions. The IDEA Center welcomes partnerships with other institutions interested in further exploring this or other questions related to the use of student ratings in assessing and improving instructional performance. Such studies are difficult to implement in real settings, but the ecological validity of the current study is one of its strengths. Second, our results were limited by the imperfect reliability of the course examinations. The restricted reliability likely diminished the magnitudes of the correlations between student ratings and performance in the course. The low correlations observed with the second exam detracted from the general pattern observed across the other four. However, the potential psychometric shortcomings of the exams developed by the instructor probably reflected the types of assessments usually administered in college courses. Third, students rated their progress on only two learning objectives that were known to be highly correlated. The outcomes were restricted to the learning of factual knowledge (e.g. terminology, classifications) and fundamental principles and theories. We defined learning as student performance on exams created by the instructor, which says nothing about incidental learning beyond what is required by course objectives. More research should be

conducted to assess the validity of student ratings of progress on other types of learning identified as relevant to the course. Fourth, student self-reported progress on learning objectives is, admittedly, an imperfect measure of how much students have actually learned in a course and, therefore, of teaching effectiveness. Many other factors, including student motivation and student ability, affect how much students learn (Cohen 1981). However, McKeachie (1979) established long ago that external factors such as individual differences, course characteristics and instructor characteristics do not strongly influence student self-reported ratings. Fifth, our findings are most relevant to instructors who clearly specify their course objectives and desirable student learning outcomes. Articulating objectives to students may have helped them to be more aware of their own learning.

Even with these limitations, this study supports the view that student self-reports of learning can be a valid indirect measure of student learning. However, any evaluation of student learning and teaching must include multiple assessments. Students are appropriate sources of evidence when they are asked to judge their own learning, the quality of student–teacher interactions, their views of the instructor's professional and ethical behaviour and grading practices, their own work habits and motivation to take the course, and the instructor's ability to communicate. However, students are not very good judges of the course content or the instructor's level of scholarship (Braskamp and Ory 1994). Other indicants of teaching effectiveness, such as ratings by colleagues, overall student ratings of the course and instructor, peer and administrator reviews of course materials, reflective statements, student products and so forth should be considered.

The current findings have implications for college instructors who want feedback from students about how much they believe they have learned relative to important and essential objectives. Most, if not all, college course syllabi list instructional objectives that signal students which course material is most important to learn. Such instructional pointers indicate to students what the instructor considers essential or important knowledge or which skills students should acquire. Asking students to rate themselves on those objectives at the end of the course makes good sense as one indicant of their achievement of learning outcomes.

Future research should replicate the current study across multiple disciplines and instructors. Student progress could also be assessed on higher-order objectives, such as applying knowledge, developing creative capacities, and learning to analyse and critically evaluate ideas. In addition, because the ultimate goal of evaluation should be improvement, researchers should continue to examine which teaching methods are most highly correlated with student progress on relevant learning objectives and exam performance. The IDEA Center examines the relationships between teaching methods and learning objectives and then recommends to instructors which teaching methods might be employed more frequently to support greater student learning for a specific objective. Further research should correlate student ratings of the instructor's use of various teaching methods with students' exam performance.

Note

1. All instruments provided by The IDEA Center are copyrighted. No part of the instruments may be used without prior written consent from the Center. Any other use will be viewed as copyright infringement and a violation of academic integrity. Violations will be pursued accordingly.

Notes on contributors

Stephen L. Benton is a senior research officer at The IDEA Center. He has published extensively in the field of educational psychology on such topics as memory for prose, writing processes, academic studying, and alcohol/drug abuse prevention. His current research interests include student ratings of instruction, teaching and learning, and higher education leadership.

Dennis Duchon is the E.J. Faulkner Professor of Management at the University of Nebraska. He has published widely in management journals in the areas of leadership, decision-making, workplace spirituality, motivation, and complexity science.

William H. Pallett is president of The IDEA Center. His current research interests include student ratings, teaching and learning, higher education leadership, and organizational transitions. He has presented at numerous conferences and campuses about the above topics.

References

Baird, J.S. 1987. Perceived learning in relation to student evaluation of university instruction. *Journal of Educational Psychology* 79: 90–1.

Bloom, B.S. 1956. *Taxonomy of educational objectives: Cognitive domain*. New York, NY: David McKay.

Braskamp, L.A., and J.C. Ory. 1994. *Assessing faculty work: Enhancing individual and institutional performance*. San Francisco, CA: Jossey-Bass.

Cashin, W.E. 1995. *Student ratings of teaching: The research revisited. IDEA Paper No. 32*. Manhattan, KS: Center for Faculty Evaluation and Development, Kansas State University.

Cohen, J. 1977. *Statistical power for the behavioral sciences.* (Rev. ed.). New York, NY: Academic Press.

Cohen, P.A. 1981. Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research* 51: 281–309.

Feldman, K.A. 1989. The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education* 33: 317–75.

Hoyt, D.P. 1973. Measurement of instructional effectiveness. *Research in Higher Education* 1: 367–78.

Hoyt, D.P., and W.E. Cashin. 1977. *IDEA Technical Report No. 1: Development of the IDEA System.* Manhattan, KS: Kansas State University.

Hoyt, D.P., Y. Chen, W.H. Pallett and A.B. Gross. 1999. *Technical Report No. 11: Revising the IDEA system for obtaining student ratings of instructors and courses.* Manhattan, KS: The IDEA Center.

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

Hoyt, D.P., and S. Perera. 2000. *IDEA Research Report No. 2*: *Validity of the IDEA student ratings of instruction system: An update.* Manhattan, KS: The IDEA Center.

Krathwohl, D.R. 1998. *Methods of educational and social science research: An integrated approach*. 2nd ed. New York, NY: Longman.

Krathwohl, D.R., B.S. Bloom, and B.B. Masia. 1964. *Taxonomy of educational objectives: Affective domain*. New York, NY: David McKay.

Marks, M., D. Fairris, and T. Beleche. in press. Do course evaluations reflect student learning? Evidence from a pre-test/post-test setting. *Review of Economics and Statistics.*

Marsh, H.W. 1984. Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology* 76: 707–54.

McKeachie, W.J. 1979. Student ratings of faculty: Reprise. *Academe* 65: 384–97.

Messick, S. 1995. Validity of psychological assessment: Validation of inferences from persons, responses, and performances as scientific inquiry into score meaning. *American Psychologist* 50: 741–9.

Ory, J.C., and K. Ryan. 2001. How do student ratings measure up to a new validity framework? In *Online student ratings of instruction: New directions for teaching and learning*, ed. T.D. Johnson and D.L. Sorenson, 27–44. San Francisco, CA: Jossey-Bass.

Remmers, H.H. 1930. To what extent do grades influence student ratings of instructor? *Journal of Educational Psychology* 21: 314–6.

Rosenthal, R. 1969. Task variations in studies of experimenter expectancy effects. *Perceptual & Motor Skills* 29: 9–10.

Appendix

The IDEA Learning Objectives[1]

(1) Gaining factual knowledge (terminology, classifications, methods, trends)
(2) Learning fundamental principles, generalisations or theories
(3) Learning to apply course material (to improve thinking, problem solving and decisions)
(4) Developing specific skills, competencies and points of view needed by professionals in the field most closely related to this course
(5) Acquiring skills in working with others as members of a team
(6) Developing creative capacities (writing, inventing, designing, performing in art, music, drama, etc.)
(7) Gaining a broader understanding and appreciation of intellectual/cultural activity (music, science, literature, etc.)
(8) Developing skill in expressing oneself orally or in writing
(9) Learning how to find and use resources for answering questions or solving problems
(10) Developing a clearer understanding of, and commitment to, personal values
(11) Learning to analyse and critically evaluate ideas, arguments and points of view
(12) Acquiring an interest in learning more by asking my own questions and seeking answers