

# Adverse Changes In Faculty Behavior Resulting From Use Of Student Evaluations Of Teaching: A Case Study

Robert E. Pritchard, Rowan University, USA

Gregory C. Potter, Rowan University, USA

## ABSTRACT

*Based on a detailed literature review and longitudinal analysis, this paper explores the possible underlying causes of the decline in the number of hours per week graduating business seniors indicated they studied during their senior year. The study was conducted at an AACSB accredited college of business at a regional university. The study indicates that the decline in hours studied was likely an unintended result of using a process designed to demonstrate continuous improvement in teaching. The process utilized the Educational Testing Service's SIR II student evaluation instrument as the only measure of teaching quality/effectiveness. The study concludes that the process may have pressured some instructors to sacrifice teaching rigor in an attempt to obtain more favorable student evaluations, thereby precipitating the decline in hours studied.*

**Keywords:** Improvement in teaching; Evaluation of teaching; Teaching effectiveness; Rigor in teaching; Student evaluation of teaching; Assessment of education; Student study time; Teaching quality

## INTRODUCTION

*A* ACSB International – The Association to Advance Collegiate Schools of Business (AACSB) requires that accredited colleges of business demonstrate continuous improvement in teaching. This paper focuses on the use of a process (Process) intended to demonstrate continuous improvement in teaching at an AACSB accredited college of business (COB) at a regional university (University).

Specifically, this paper examines the possible relationship between 1) the COB's use of the Process and 2) the decline in the number of hours per week COB graduating seniors indicated that they studied during their senior year. The Process utilized the Educational Testing Service's SIR II student evaluation instrument as the only measure of teaching quality/effectiveness. This instrument is one of a number of instruments categorized in the literature as "student evaluations of teaching" (SETs).

SETs are student feedback instruments used to obtain course and student evaluations. Most SETs incorporate a series of Likert-like rating scales permitting students to rate various aspects of the course/instruction using a number of choices that might, for example, range from "very effective" to "very ineffective."

## LITERATURE REVIEW

SETs are used widely in higher education. Nonetheless, questions about their reliability, validity, and appropriateness for measuring effective teaching have persisted since their inception (Ahmadi & Cotton, 1998; Algozzine et al, 2004; Baldwin & Blattner, 2003; Wolfer & Johnson, 2003). For example, and especially relevant to this research paper, is the impact and possible bias of various student SET responses in formal faculty evaluation based on factors such as time spent on course study, class attendance policies, and involvement in evaluative activities. The relationships between student SET responses and factors such as these appear to be ambiguous, if not largely uninvestigated (Burns & Ludlow, 2005; Nonis & Hudson, 2006; Stinebrickner & Stinebricker, 2004).

A predominant theme in this literature concerns the use of evaluations as “customer satisfaction” surveys (Smith, 2004), with attendant concerns about the “leniency hypothesis” (faculty will receive better evaluations if they lower performance expectations) (Gump, 2007; Heckert, Latier, Ringwald & Silvey, 2006). Such behavior, at the extreme, can lead to what Zorn (Academic Leader, 2005) calls the “Imposter Phenomenon” (faculty who feel “fraudulent” because perceptions they hold for themselves do not align with what they believe the academic culture is telling them).

Simpson and Siguaw (2000) acknowledge that SETs are critical to the economic and psychological well-being of faculty. Faculty who genuinely are committed to good teaching practices may, however, be subject to hurtful or malicious ratings; consequently, unethical faculty behavior to influence ratings favorably may result, leading to a destruction of educational objectives.

Trout (2000) remarks that numerical forms, used typically to reward or punish classroom behavior, encourage instructors to “dumb down” their teaching. He comments that teacher evaluations may create an incentive for teachers to do the wrong thing, namely, to please students rather than teach them. In the “consumerist academy,” as Titus (2008) calls it, instructors, as Trout suggests, are pressured to compete with one another for tenure, promotion and perks which could lead to sacrificing teaching integrity.

Rigor in teaching is also discussed by Emery et al (2003). They remark that lecturers who perceive performance appraisals as popularity contests will treat their students as customers, reflecting a “self-interest” and “self-preservation” that may supplant rigorous teaching. Felton et al (2004) note that instructors offering easy courses tend to be rated more highly.

Relevant to both the “Dynamics of Teacher-Student Interaction” and “Classroom Setting” aspects of student evaluation analysis of this case study, Yunker and Yunker (2003) note that SET scores are consistently higher (for the same instructor) in upper-division and graduate courses than for lower-division, introductory courses. These researchers also comment that students tend to view required courses in accounting, economics, and statistics as obstacles rather than stepping stones to success.

To summarize, the literature indicates that the use of SETs may pressure faculty to “dumb down” their teaching. Such pressure may be especially prevalent in required lower-division courses.

## **DESCRIPTION OF THE PROCESS INTENDED TO DEMONSTRATE CONTINUOUS IMPROVEMENT IN TEACHING**

### **Underlying Assumption**

The Process described below was based on the underlying assumption that improvement in the average (mean) of COB faculty SIR II evaluations over time would indicate that teaching in the COB had improved, thereby demonstrating continuous improvement in teaching. Although intuitively appealing, this underlying assumption may well have been erroneous for two reasons.

First, the assumption begs definition of “improvement in teaching.” One can argue that the best measure of improvement in teaching is increased student learning. If that argument is accepted, then the demonstration of improvement in teaching depends on demonstrating increased learning. That is, if teaching is improving, then *ceteris paribus*, students should be learning more.

Only two of the SIR II evaluative categories, Categories F and I, focus on measuring various aspects of students’ perceptions of their learning. All of the other categories focus on student perceptions of the instructors’ performance (e.g., instructor ability to organize and plan the course, instructor ability to communicate, etc.), the usefulness of supplemental instructional methods, student effort and involvement, and course difficulty, workload and pace.

Category F (Course Outcomes) includes the following five SIR II items all of which focus on the students' perception of how much they learned. Students respond to each item using a five point Likert-like scale ranging from "very effective" to "ineffective."

- 29. My learning increased in this course.
- 30. I made progress toward achieving course objectives.
- 31. My interest in the subject area has increased.
- 32. This course helped me to think independently about the subject matter.
- 33. This course actively involved me in what I was learning.

Category I (Overall Evaluation) includes only one item.

- 40. Rate the quality of instruction in this course as it contributed to your learning.

Given the fact that the large majority of SIR II items are not focused on student learning, it may be erroneous to assume that a longitudinal improvement in the average COB faculty SIR II ratings equates to improvement in teaching.

Second, suppose that the items in the SIR II (or any other SET) could provide valid (effective) measures of teaching effectiveness. Even so, based on the literature review, it is obvious that student responses might well be biased based on a number of issues including course difficulty, expected work requirements and the like. Therefore, although the instrument may provide for valid measures of teaching effectiveness, the evaluative results may well be unreliable because they are biased.

Consequently, one can conclude that any longitudinal improvement in mean COB faculty SIR II ratings does not necessarily indicate improvement in teaching. That is, the underlying assumption may be erroneous.

### **Details of the Process**

For many years the COB used the Educational Testing Service's (ETS) SIR (and more recently the SIR II) student evaluation instrument along with peer evaluations as the measures of teaching effectiveness. Given the faculty's and administration's experience using the SIR II, they decided to develop a process designed to demonstrate continuous improvement in teaching. The process employed the SIR II as the single measure of teaching effectiveness and did not include peer evaluations.

The process was implemented spring 1999 but lacked useful benchmarks for measuring continuous improvement in teaching. Starting with the fall 2001 semester, the Educational Testing Service (ETS) provided the COB with a "Combined Report University College of Business" (Combined Report). Each semester's Combined Report provides the COB averages of student evaluations by SIR II category for seven <sup>1</sup> of the SIR II categories. The seven categories include the following:

- A. Course Organization and Planning
- B. Communications
- C. Faculty/Student Interaction
- D. Assignments, Exams, and Grading
- F. Course Outcomes
- G. Student Effort and Involvement
- H. Overall Evaluation

Each semester a copy of the Combined Report was provided to each participating COB faculty member along with the SIR II reports for her/his classes. As such, the Combined Reports provided each COB faculty member with a set of seven benchmarks that she/he could use to compare her/his student evaluations for a given semester with the averages for the COB as a whole.

Once the Combined Reports became available, COB faculty members were required to prepare annual teaching improvement plans focusing particular attention on those SIR II categories for which their individual results were lower than the COB Combined Report averages. The underlying argument for requiring each faculty member to focus on her/his evaluations that were below the COB Combined Report averages was to try to boost the COB combined averages.

Faculty members were required to submit to their chairperson and the COB dean copies of their SIR II results for all of their classes each semester, as well as a teaching improvement plan, at the end of the academic year. Thus, the SIR II results for all COB faculty members (including probationary faculty members and faculty members applying for promotion) were reviewed and critiqued by the COB dean, as well as by each faculty member’s department chair for every class, every semester.

The requirement for all COB faculty members to submit to their chairperson and the COB Dean copies of their SIR II results for all of their classes each semester defeated the self-selecting methodology for reporting student evaluations delineated in the University-Union agreements pertaining to recontracting, tenure and promotion. As a part of the recontracting and tenure and promotion processes, the self-selecting methodology permits a faculty member to select and submit for review only a limited number of her/his SIR II reports (generally the most favorable). The Process requirement of having to submit copies of all SIR II reports to the chairperson and COB Dean placed added pressure on probationary faculty and promotion applicants to obtain superior student evaluations.

Finally, if a faculty member refused to participate in the Process, funds needed to support scholarly activities (such as travel to conferences, research assistance, computer hardware and software, etc., normally amounting to \$3,000 per year) were denied.

**MEASURING THE CHANGE IN THE NUMBER OF HOURS PER WEEK THAT COB GRADUATING SENIORS INDICATED THEY STUDIED**

In examining the possible untoward relationship between the COB’s use of the SIR II and the decrease in the number of hours per week graduating seniors indicated that they studied, it is first necessary to describe the Process.

Commencing with its graduating class spring semester 1998, the COB has used the EBI Undergraduate Business Exit Assessment <sup>2</sup> (Assessment). EBI is the acronym for Educational Benchmarking Inc (www.webebi.com). The Assessment has been administered to those seniors graduating in the spring every two years since 1998. The Assessment utilizes a survey instrument that is quite broad and includes questions relating to various student demographics as well as questions covering 14 factors relating to faculty, instruction of business courses, University services, etc. One of the demographics included is Categorical Question D006: “Average number of hours studied per week during the past academic year.”

The level of COB student engagement in their studies as indicated by graduating COB student responses in the Assessment is shown in Table I. The ranges of hours studied per week shown in Table 1 are the same as the response categories for Categorical Question D006.

**Table 1<sup>3</sup>**  
**COB Graduating Senior Study Data 1998 through 2006: Percentages of Students Indicating “Average Number of Hours Studied per Week during the Past Academic Year”**

Average Hours Studied per Week	2006 n = 219 students	2004 n = 171 students	2002 n = 184 students	2000 n = 195 students	1998 n = 114 students
0 - 5	33%	23%	33%	25%	19%
6 - 10	31	36	32	29	27
11 - 15	23	22	22	24	25
16 - 20	6	10	6	11	18
21 - 25	6	5	6	6	7
26 - 30	0	2	1	2	3
> 30	0	2	1	3	2

A review of the data in Table 1 indicates that the number of hours per week COB graduating seniors indicated that they studied decreased significantly since 1998, albeit there have been variations from year to year. Especially concerning is the increase in the percentage of students who indicated that they studied five hours or less per week. This increased to 33 percent in 2006 versus 19 percent in the baseline year, 1998. Similarly, the percentage of students indicating that they studied 10 hours or less increased to over 64 percent in 2006 versus 46 percent in 1998.

The University is designated a “Master’s I” Carnegie Class<sup>4</sup> institution. The statisticians at EBI analyzed the COB’s “Average Hours Studied per Week” data for 1998 and compared it to the 1998 EBI data for the “Master’s I” institutions included in EBI’s data base. They concluded that for 1998 there was no statistically significant difference in the COB’s data and the “Master’s I” institutions’ data pertaining to “Average Hours Studied per Week.” By contrast, after analyzing the 2006 data, the statisticians concluded that the COB’s “Average Hours Studied per Week” were statistically significantly lower than the comparable “Master’s I” institutions.

Finally, it should be noted that during the period from 1998 to 2006, the percentage of full-time versus part-time COB seniors enrolled during the spring semester increased from approximately 61 percent (of n = 415 students) in 1998 to 75 percent (of n = 371 students) in 2006. Concurrently, as expected with a higher percentage of full-time COB seniors, the mean attempted semester hours for COB seniors increased from 11.16 spring 1998 to 12.55 spring 2006. The increase in the percentage of full-time COB seniors (and the concurrent mean attempted semester hours) could be expected to result in an increase in the number of “Average Hours Studied per Week;” it did not. Thus, not only was there a significant decrease in “Average Hours Studied per Week” but this decrease took place during a period when the mean semester hours attempted by COB seniors increased.

## **ANALYSIS AND CONCLUSIONS**

The Process of using SIR II student evaluations in an attempt to demonstrate continuous improvement in teaching placed COB faculty members – and especially those who were probationary or applying for promotion – in the position of having to constantly compete with one another to meet or exceed the COB Combined Report averages in seven SIR II categories. Regardless of the effort each faculty member exerted in attempting to improve her/his teaching, on average at least half of all COB faculty members received one or more below average ratings each semester.

Based on the literature, it can be hypothesized that some COB faculty members may have bartered their high educational standards for better student evaluations. They may have felt pressure to sacrifice academic rigor, perhaps succumbing to ingratiating behaviors designed to obtain good course evaluations rather than focusing on the material students needed to master. Such a sacrifice in rigor could explain the decrease in the number of hours per week COB graduating seniors indicated that they studied. Brent and Felder (1999) note that faculty, especially at the beginning of a course, have a challenging role both as “gatekeeper” and “coach” (ensuring that students learn what is needed for professional success and encouraging students to surpass hurdles faculty may have set for them, respectively). To help motivate and challenge students, these researchers suggest that faculty need to provide assignments that align with clearly articulated course goals and objectives.

Unfortunately, as noted in the Literature Review, the literature relating to student evaluations and contextual variables (such as grading leniency and course expectations) indicates that these variables influence students’ evaluations. Emery et al, (2003) and Felton et al (2004), for example, note that faculty members using student evaluation instruments may be encouraged to teach less rigorously. Based on the decrease in hours graduating seniors indicated that they studied per week, it would appear that in this instance the pressure to obtain good SIR II evaluations outweighed the goal of challenging and motivating students to meet their full potentials.

The Process used by the COB, in a well-intentioned attempt to demonstrate continuous improvement in teaching, may have been either the primary cause for or contributed to the steep decrease in the number of hours graduating seniors indicated that they studied.

Alternatively, it can be hypothesized that the Process actually prompted COB faculty members to improve their teaching so that students learned as much or more with fewer hours of study. Similarly, it can also be hypothesized that the quality of the students improved over the years and, therefore, they required fewer hours of study.

Examining these two alternative hypotheses individually or collectively necessitates answering a critical question: “what is the primary goal of higher education?” If the answer is preparing each student to achieve her/his full potential, then, given the decrease in the number of hours studied, one could conclude that the COB faculty did not achieve the primary goal. If quality of instruction and/or quality of students had increased, the faculty members could have raised expectations and challenged students to study more, not less. This conclusion, of course, is based on the premise that greater student participation in study activities leads to more learning and more learning better prepares students to reach their full potentials.

Another factor warrants consideration. Was there any statistically significant change in the SIR II results during the period of the study? (Citation to be added) provided a detailed statistical analysis of the COB’s Combined Report SIR II results on a semester-by-semester basis commencing fall 2001 through spring 2007. They concluded that during this seven-year period, there was no statistically significant improvement in any of the seven SIR II categories included in the Combined Report.

So, tragically, even if the Process’s underlying assumption, that “improvement in the average (mean) of COB faculty SIR II evaluations over time would indicate that teaching had improved, thereby demonstrating continuous improvement in teaching,” was correct, the Process did not demonstrate any improvement in teaching.

Of course, other factors may have contributed to the decline in the number of hours graduating seniors indicated they studied. Nonetheless, as the Literature Review indicates, instructors may sacrifice rigor or take other actions in an attempt to obtain more favorable student evaluations. Ironically, even if the Process only contributed to the decrease in the number of hours COB graduating seniors studied, application of the Process likely resulted in fewer COB graduates being prepared to meet their full potentials.

#### **AUTHOR INFORMATION**

**Robert Pritchard** completed both his undergraduate degree in physics and MBA at Drexel University, his MA in applied economics at Wharton, and his doctorate in Education Administration at the University of Pennsylvania. He has authored/co-authored nine books in the fields of finance, small business management and marketing as well as written over 250 trade journal articles. He has consulted and provided financial training for many businesses and trade associations throughout the US. His research interests include real estate, personal financial management, retirement planning, and Social Security. He specializes in applied financial research and pedagogical research principally pertaining to the teaching/learning processes in business and finance.

**Gregory Potter** is the Associate Dean of Library Services at Rowan University, where he coordinates library instruction and information literacy programs. He has collaborated with Rowan University faculty in published research on academic assessment, pedagogy, and career development programs. Dr. Potter earned his doctorate in educational foundations at Rutgers University, a Master of Science in Library Science at Villanova University, a Master of Arts in Public Relations at Rowan University, and a Bachelor of Arts in English from the University of Pennsylvania.

#### **FOOTNOTES**

1. Although there are other categories within the SIR II evaluation instrument (such as course difficulty, work load, and pace), ETS does not compute the averages for these.
2. The EBI Undergraduate Business Exit Assessment was previously called the AACSB/EBI Undergraduate Business Exit Study.
3. The EBI Undergraduate Business Exit Assessment was administered spring 2008. Unfortunately, the survey instruments were not distributed completely; a disproportionate 75 percent of the respondents were

students majoring in accounting and finance. Therefore, the results were not included in Table 1. Nonetheless, the statisticians at EBI compared the 2008 results with the 2006 results and determined that there was no statistically significant change from 2006 to 2008.

4. This means the institution offers a wide range of baccalaureate programs and is committed to graduate education through the master's degree and the institution confers 40 or more master's degrees annually in three or more disciplines.

## REFERENCES

1. Academic Leader, (2005). Academic culture feeds the imposter phenomenon. *Academic Leader*, 21(8), 7-8.
2. Ahmadi, R., & Cotton, E. (1998). Assessing students' ratings of faculty. *Assessment Update*, 10(5), 6-7.
3. Algozzine, J., Beattie, J., Bray, M., Flowers, C., Gretes, J., Howley, L., Ganesh, M. & Spooner, F. (2004). Student evaluation of college teaching. *College Teaching*, 52(4), 134-41.
5. Baldwin, T., & Blattner, N. (2003). Guarding against potential bias in student evaluations: What every faculty member needs to know. *College Teaching*, 51(1), 27-32.
6. Brent, R. & Felder, R. (1999). It's a Start. *College Teaching*, 47(1), 14-17.
7. Burns, S., & Ludlow, L. (2006). Understanding student evaluations of teaching quality: the contributions of class attendance. *Journal of Personnel Evaluation Education*, 18, 127-138.
8. Emery, C., Kramer, T., & Tian, R. (2003). Return to academic standards: a critique of student evaluations of teaching effectiveness. *Quality Assurance in Education*, 11(1), 37-46.
9. Felton, J., Mitchell, J., & Stinson, M. Web-based student evaluations of professors: the relations between perceived quality, easiness and sexiness. *Assessment & Evaluation in Higher Education*, 29 (1), 91-108.
10. Gump, S. (2007). Student evaluations of teaching effectiveness and the leniency hypothesis: a literature review. *Educational Research Quarterly*, 30(3), 55-68.
11. Heckert, T., Latier, A., Ringwald, A., & Silvey, B. (2006). Relation of course, instructor, and student characteristics to dimensions of student ratings of teaching effectiveness. *College Student Journal*, 40 (1), 195-203.
12. Nonis, S., & Hudson, G. (2006). Academic performance of college students: Influence of time spent studying and working. *Journal of Education for Business*, 81(3), 151-159.
13. Simpson, P., & Siguaw, J. (2000). Student evaluations of teaching: an exploratory study of the faculty response. *Journal of Marketing Education*, 22(3), 199-201.
14. Smith, G. (2004). Assessment strategies: what is being measured in student course evaluations? *Accounting Education*, 13(1), 3-28.
15. Stinebrickner, R., & Stinebrickner, T. (2004). Time-use and college outcomes. *Journal of Econometrics*, 121, 243-269.
16. Titus, J. (2008). Student ratings in a consumerist academy: leveraging pedagogical control and authority. *Sociological Perspectives*, 51(2), 398-422.
17. Trout, P. (2000). Flunking the test: the dismal record of student evaluations. *Academe*, 86(4), 58-62.
18. Trout, P. (2000). Teacher evaluations. *Commonweal*, 127(8), 10-12.
19. Wolfer, T., & Johnson, M. (2003). Re-evaluating student evaluation of teaching; the teaching evaluation form. *Journal of Social Work Education*, 39(1), 111-121.
20. Yunker, P., & Yunker, J. (2003). Are student evaluations of teaching valid? Evidence from an analytical business core course. *Journal of Education for Business*, 78(6), 313-317.

**NOTES**