International Journal of Teaching and Learning in Higher Education 2011, Volume 23, Number 2, 226-235
http://www.isetl.org/ijtlhe/ ISSN 1812-9129

Student Evaluation of Teaching: An Instrument and a Development Process

Kumar Alok
Chandragupt Institute of Management Patna

This article describes the process of faculty-led development of a student evaluation of teaching instrument at Centurion School of Rural Enterprise Management, a management institute in India. The instrument was to focus on teacher behaviors that students get an opportunity to observe. Teachers and students jointly contributed a number of desirable and undesirable performance examples that went through a process of filtration using mean-difference item response analysis and factor analysis. The final instrument has 18 examples to be rated on a six-point scale. It was used with a formative focus; however, the post-implementation experiences indicated the need for limited summative focus as well. New students need to be educated about student evaluation of teaching and its relevance for a quality academic life. It also emphasizes the need for open communication and a climate of trust for a successful student evaluation of teaching.

Introduction

Student evaluation of teaching (SET) remains a keenly debated issue (Langbein, 2005; Murray, 2005). It is one of the most criticized (Ellis, Burke, Lomire & McCormack, 2003; Wright, 2006) and yet the most prevalent (Richardson, 2005; Shevlin, Banyard, Davies, & Griffiths, 2000) practices in higher education. Most US business schools use SET to determine teaching effectiveness (Comm & Mathaisel, 1998). The All India Council for Technical Education (AICTE), the major regulator for management education in India, considers SET as an important indicator of the academic quality of a management institute. Therefore, it becomes important for a new college to institute such a practice in India.

This article focuses on SET instrument developed for Centurion School of Rural Enterprise Management (CSREM), India and its systemic usage. The instrument uses 18 behavioral performance examples along three performance dimensions on a six-point rating scale. It marries the simplicity of a graphic rating scale with the precision of critical incidents. It captures 10 key areas of faculty performance that students can observe: course design, instruction skills, depth of knowledge, facilitation skills, student-faculty interaction, ability to motivate, quality of assignments, organization of assessment, perceived fairness, and quality of feedback.

The article begins by reviewing SET literature and discussing the context in which the instrument was developed. It goes on to describe in detail the procedures followed and discusses the reactions of students and teachers. It discusses the post-implementation issues and proposed counter-measures, and it concludes by pointing out the limitations as well as the insights of the study.

Literature Review

The North American model is the dominant model of management education (Clegg & Ross-Smith 2003). It emphasizes analytical generalizations for developing broad knowledge concerning business functions such as marketing, finance, human resource management, production and operations, and systems. Today it is facing tremendous demands for relevance and accountability (Gosling & Mintzberg, 2006; Knowles & Hensher, 2005). Rising salaries of management graduates, rising costs of management education, and media rankings have fuelled these demands (Zell, 2001). Business schools are struggling to meet the demands of a fast changing business world (Knowles & Hensher, 2005).

Leading business schools are responding through radical curricular reforms (Bisoux, 2007). The new focus is on flexibility and an integrated view of business. Issues of practical relevance such as leadership, globalization, communication skills, problem identification in ambiguous situations, and self awareness are going to anchor management education for the 21st century (HBS Centennial Colloquia Report, 2009). Teaching effectiveness is critical for the success of any such initiative.

While other measures of teacher performance have gained momentum, SET continues to hold its primacy (Arreola, 2007; Emery, Kramer & Tian, 2003; Murphy, Maclaren, & Flynn, 2009; Seldin, 2006). It largely reflects the effect of a teacher (Marsh, 1982; Marsh & Roche, 1997; Wright, 2006). Besides, it considers feedback from students who are best placed to observe the in-class performance of a teacher.

The validity of SET information has been a contentious issue (Berk, 2005). SET scores do not necessarily correlate with overall teacher evaluation that typically includes peer and supervisor evaluations as well (Dunegan & Hrivnak, 2003). Its methods are

not necessarily in sync with the transformed technological environment of a management classroom (Serva & Fuller, 2004). On the other hand, SET scores are known to reflect students' perception of the teacher's attitude, presentation skills, reliability, and learning skills (Kim, Damewood & Hodge, 2000). These factors are amenable to learning and improvement.

SET can be used for formative or developmental and summative or administrative purposes (Murphy, Maclaren, & Flynn, 2009; Theall & Franklin, 1990), though considerable disagreement exists on the issue of choosing one over the other (Centra, 1993; Miller, 1987; Seldin, 1984; Waller, 2004; Younes, 2003). Arreola (2007) warned that teachers would consider any attempt to evaluate as punitive unless it is linked with professional development opportunities. In other words, a summative purpose would be counterproductive in the absence of a well articulated formative purpose. Centra (1993) suggested that teacher evaluations should be formative to begin with. It would help teachers understand what is required of them before being judged in a summative manner. In any case, SET should not be considered as sufficient for summative decisions (Berk, 2005).

SET operates within the limits of rationality (Waller, 2004). Rationality is often associated with an emphasis on objectivity in evaluation, which explains the quantitative nature of most SET instruments. Qualitative feedback can further enrich the outcome of student evaluations (McKone, 1999).

The SET literature emphasizes the multidimensionality of teaching (Arreola, 2007; Braskamp & Ory, 1994; Centra, 1993; Feldman, 1988; Fink, 2008; Marsh, 1984; 1993). Most SET forms are multidimensional in nature as well (Marsh & Dunkin, 1992). Overlapping of dimensions is not known to significantly affect student ratings (Marsh, 1987).

Scholars are increasingly emphasizing that the reliability, validity, and usefulness of a SET instrument should be determined at the institution where it is to be used (L'Hommedieu, Menges, & Brinko, 1990; Murray, Rushton, & Paunonen, 1990; Seldin & Angelo, 1997). Harrington and Schibik (2003) reported that despite the availability of commercially available instruments, more than 80% of the surveyed institutions used "home grown" instruments to address faculty preferences. Quality of teaching itself is a discipline-specific construct as disciplinary differences affect beliefs about the nature of knowledge and learning, teaching practices, and perceptions about what is effective teaching and how to evaluate it (Braxton & Hargens, 1996; Cashin, 1990, 1995; Hutchings & Shulman, 1999). Therefore many scholars have advocated for discipline and culture-specific faculty evaluation systems (Aubrecht, 1984; Cashin, 1990; Geis, 1984). Institution-specific SET instruments are in a better position to address many of such concerns.

Perceptions of teachers and administrators are often at variance when it comes to evaluation (Younes, 2003). SET ratings have considerable influence over administrative decisions (Emery et al., 2003), whereas, they often evoke cynicism and even hostility among teachers (Franklin & Theall, 1989; Nasser & Fresko, 2002). Cashin (1999) argued that teachers would use the evaluation data provided they have confidence in its reliability and validity. Franklin and Theall (1989) found that teachers with greater awareness about the research on student evaluations showed more positive attitude toward their usage. Involving teachers in the development of SET instrument improves the chances of its acceptance (Seldin & Angelo, 1997).

Background

India has more or less adopted the North American model of business education though with a difference. Indian business schools are largely driven by the prospects of lucrative placements for the students. Recruiting organizations tend to approach newly established business schools to meet their basic needs. Thus a newly established business school may choose to cater to niches such as telecom, retail, and insurance, or they may offer a traditional commoditized business education. A commoditized business education involves less operational costs as teachers would be readily available.

Offering the AICTE-approved two-year traditional management program was a strategic choice for CSREM. In June 2006, CSREM was established in Paralakhemundi in the state of Odisha, India. It is the result of a public-private partnership among the Government of India, the Government of Odisha, and the CSREM Trust. Being an autonomous institution, it allows teachers to design their courses, pedagogy, and assessment within the limits of the institutional framework. It can admit 120 students per year, though the actual intake so far has remained close to 70. There are 14 permanent teachers and a number of visiting professors for the program.

The AICTE mandates SET for all credit courses taught in an approved management program in India. Therefore, SET came early on the administration's agenda. It was decided to develop an institution-specific SET instrument with a formative focus and a limited summative significance. It was to be a quantitative instrument with a place to write comments. The administration asked for faculty volunteers to lead the instrument development. The author volunteered to take the project from concept to commission. From the beginning, students and teachers were apprehensive yet enthusiastic about the instrument development process.

The Instrument Development Process

The first task was to decide what to evaluate. One can choose to evaluate performance on three bases: traits, behaviors, and outcomes (Mello, 2004). Traits are largely stable and hence not suitable for developmental purposes. Learning outcomes, though they correlate well with student ratings (Theall & Franklin, 2001), are a function of student characteristics such as ability, attitude and motivation, institution characteristics such as class size, learning resources and institutional climate, and teaching effectiveness (Berk, 2005). Teaching effectiveness alone cannot account for learning outcomes. Teacher efforts can be better appreciated by evaluating their job-related behaviors. Performance examples that students can directly observe are well suited for SET purposes.

To begin, the author studied about a dozen of SET instruments used by various Indian business schools. Because the focus was to develop an institution-specific SET instrument, other instruments were studied primarily to gain insights into what is typically assessed. This study was supplemented with an exhaustive literature review. The author made a brief presentation about the literature and argued for the behavioral basis of SET that convinced the teachers and the administrators to accept it. Identifying the performance dimensions for evaluation was next on the agenda.

A brainstorming session with students and teachers resulted in the identification of 16 key areas along three performance dimensions as given in the Table 1. The key areas identified were more or less in sync with the evaluation factors commonly reported in the literature (Braskamp & Ory, 1994; Centra, 1993; Feldman, 1988). Further, the students wanted to keep the instrument short and less demanding on time.

Research about the relative effectiveness of behaviorally-anchored rating scales and Likert-type graphic rating scales in the context of SET is more or less inconclusive (Cook, 1989; Eley & Stecher, 1997). Agree/disagree type Likert scale is less demanding on time as compared to BARS; however, a BARS-type critical incident-based performance example offers precision. The author decided to construct a Likert-type graphic rating scale using precise performance examples.

A workshop was conducted to train students and teachers in writing effective and ineffective examples along the performance dimensions and the key areas identified earlier. To begin with, the author presented a number of written examples of effective and ineffective performance for deliberations. The participants were encouraged to critique the substantive as well as the formal aspects of the examples. They developed a few sample examples and presented them for critique. By the end of the workshop, they were assigned the key areas and asked to submit four examples, two each on effective and ineffective performances. Within a week, they submitted 116 examples. It was time to screen unacceptable examples.

Example Selection Criteria

Apart from the frequency of mention, five other criteria were used to select examples:

1. Examples must be observable. Idealistic or non-observable examples were eliminated.
2. Examples must describe the teaching performance. Examples describing administrative or environmental aspects were eliminated, e.g. “This teacher used to take classes in the evening.”
3. Examples must not be biased toward a particular sex, caste, or state.
4. Examples must not be offensive in nature. Examples with a potential to hurt teachers were eliminated.
5. Examples must be clear, unambiguous, and one-dimensional in meaning. Multidimensional examples were eliminated, e.g. “This teacher hardly motivated students and answered their queries.”

A total of 68 examples passed the screening criteria: 16 for course organization, 32 for quality of teaching, and 20 for assessment and feedback. Equal distribution of effective and ineffective examples for each key area was not insured at this stage. The selected examples were retranslated to ensure clarity and brevity.

Another workshop was held to assign the re-translated examples to the relevant key areas. It was required for communicating the SET results to the teachers. An example was assigned to a key area if at least 80% of the participants favored it.

Scale Construction

Forty-three students volunteered to participate in the scale construction process. They were asked to rate 68 examples on a seven-point semantic differential scale along a “worst performance – best performance” continuum. The seven-point scale was used because the respondents’ ability to reliably distinguish between adjacent categories is known to suffer with more rating points (Krosnick & Fabrigar, 1997).

It was important to identify examples invoking highly biased responses as they cannot provide any

Table1
First Identified Performance Dimensions and Key Areas

Performance Dimension	Sl. No.	Key Areas
Course Organization	1	Course Design
	2	Quality of Course Materials
	3	Course Difficulty
	4	Articulated Pedagogical Diversity
	5	Respecting the Timeframe
Quality of Teaching	6	Instructional Skills
	7	Depth of Knowledge
	8	Facilitation Skills
	9	Student-Faculty Interaction
	10	Ability to Motivate
	11	Tolerance of Disagreement
	12	Maintaining Class Discipline
Assessment & Feedback	13	Quality of Assignments
	14	Organization of Assessment
	15	Perceived Fairness
	16	Quality of Feedback

significant insight into the rater's mind. Twenty-nine such examples with the values of either Z (Skew) or Z (Kurtosis) higher than 1.96 were weeded out through an exploratory data analysis.

Responses utilizing the middle values of a scale are ambiguous to interpret. Quartile analysis helps to identify respondents with highly favorable (fourth quartile) or highly unfavorable (first quartile) attitudes toward the scale items. Twenty-two such respondents were identified at this stage.

If an example fails to discriminate between a highly favorable and a highly unfavorable respondent, then its ability to provide any useful insight is questionable. Statistically, in more than 5% cases, the difference in means between the fourth and the first quartiles for such examples can be attributed to chance. Fourteen such examples were eliminated through a mean-difference item response analysis using a two-tailed t-test.

Reliability Analysis

A Cronbach Alpha reliability analysis was conducted on the remaining 25 examples. It led to the elimination of seven items having less than 60% correlation with the scale. The remaining 18 examples showed a very good item-scale correlation with the reliability coefficient Alpha being 0.9507. Deleting any more items would have reduced the overall reliability of the scale. These 18 examples still represented the three performance dimensions; however, they could represent only 10 out of 16 key areas initially identified.

Validity Analysis

A preliminary principal component analysis gave the scree plot with a slope reducing greatly at the level of the second factor. Accordingly, a second principal component analysis was conducted for two factors. Factor loadings with less than 0.4 absolute values were suppressed to assist interpretation. Varimax rotation was used to clearly reflect the loading of different variables on either of the factors. It resulted in two factors with nine variables each as presented in the Table 2. The factors seemed to indicate the orientation of the teachers. Accordingly they were named "learning orientation" and "learner orientation"; the first factor indicated the teachers' concern for the students' learning, whereas, the second factor indicated their concern for the students. The two factors accounted for 64.269% variance.

It is known that some teachers are rated relatively high as instructors but relatively low as producers of study and learning, and vice versa (Stapleton & Murkison, 2001). Accordingly it is assumed that learning and learner orientations are two independent factors. It would be possible to score high in both or low in both or high in one and low in the other.

Face validity and criteria validity were established through the process of developing and selecting the examples themselves. The initial constructs of the three performance dimensions were subsumed under the two larger constructs: learning orientation and learner orientation.

Table 2
Factors Underlying the Performance Examples

Factor	Sl. No.	Performance Examples
Learning Orientation	1	Gave assignments that were helpful in understanding the subject better.
	2	Strictly adhered to the deadlines of assignment submission.
	3	Used to create a threatening environment in the class.
	4	Used to look confused while teaching complex topics.
	5	Used to briefly summarize the previous lecture at the beginning of each class.
	6	Never made any attempt to make the class interesting.
	7	Described the concepts and processes related to the topic with the fundamental logic behind them.
	8	Used to mention areas of improvement and the ways to improve while giving feedback to students.
	9	Emphasized only the theoretical aspect of the subject.
Learner Orientation	1	Encouraged students to think and to question.
	2	Asked students for suggestions regarding the course outline.
	3	Used to answer students' questions clearly.
	4	Often said, "I have explained the topic. It is your problem if you have not understood it."
	5	Used to take very interactive sessions.
	6	Provided course outline having helpful suggestions regarding recommended books/websites, group formation, projects, evaluation pattern and general rules for the course.
	7	Offered to explain questions and their answers once exams were over.
	8	Clearly explained the evaluation criteria to students.
	9	Encouraged students to seek his or her help whenever in need.

Pilot Test

The tendency of respondents to avoid end points of rating scales or contraction bias is widely reported in literature (Tourangeau, Rips, & Rasinski, 2000). Satisficing, i.e., the tendency of respondents to use the path of least cognitive work while responding to surveys is also well known (Krosnick & Alwin, 1987). No opinion options such as "Can't Say" might invoke satisficing, thereby effectively precluding some meaningful opinions (Krosnick et al., 2002). Contraction bias can be minimized by increasing the number of rating points while avoiding a mid-point on the scale. Satisficing can be possibly tackled by avoiding no-opinion responses. Considering these issues, the author opted to pilot the SET instrument with a forced choice six-point Likert-type graphic rating scale ranging from "Fully Disagree (FD)" to "Fully Agree (FA)" with the mid-point split into "Slightly Disagree (SD)" and "Slightly Agree (SA)." Numbers were replaced with letter codes to forestall any role that they might play in making the raters lenient.

A typical response on a particular example would more accurately represent the view of the class than a response adjusted for extremities. On the other hand, justice demands that exceptional performances should also be considered while making a statement about the overall performance of a teacher. Accordingly, the median was used to indicate the typical performance on an individual example, whereas the average of medians was used to indicate the overall performance on the two factors, the three performance dimensions and the 10 key areas. Semi-interquartile range (SIR) was used to indicate the nature of opinions on individual examples. SIR is a measure of spread or dispersion that is little

affected by extreme scores. It is for the median what standard deviation is for the mean. A SIR of 0.5 or less was taken to indicate consensus.

Students enthusiastically participated in the pilot test, and teachers eagerly waited for the results. A focus group discussion with the students and the teachers showed the general acceptance of the instrument.

The Final Instrument

The final instrument has 18 examples to be rated on the six-point scale as indicated in the Appendix. It also provides space to let students write qualitative feedback. The instrument is being implemented using intranet.

Implementation and Impact

The reliability of a SET instrument might suffer in case less than 10 students respond (Cashin, 1988). CSREM has tackled this issue to a certain extent by making an institutional policy that prohibits elective courses with less than 10 registered students. The impact of class size on SET score remains keenly debated (Fernandez, Mateo, & Muniz, 1998; Lesser & Ferrand, 2000; Marsh & Roche, 1997; Mateo & Fernandez, 1996) though, at CSREM teachers generally score high in elective courses with smaller batch sizes.

The SET instrument is in use since 2007. It has been used over 100 times for various courses. The teachers feel that the precise performance examples facilitate their understanding of the areas of improvement. The administration has shown a positive attitude toward the SET results. Consistent with the good practices recommended in literature (Arreola, 2007; Centra, 1993), the institute sponsored two teachers for attending national level faculty development programs conducted by the prestigious Indian Institutes of Management. In 2008, regular internal faculty development programs were initiated. External experts were involved in validating the course outlines and course materials. Moreover, a series of curriculum development workshops ensued in the first half of 2009.

The formative focus went well with the teachers for sometime before they started feeling the need to get recognized. In faculty meetings, the issue of recognition was often raised. In their view, the instrument succeeded in measuring their orientation as well as areas of improvement, but failed in discriminating between excellent teachers and good teachers. A certain degree of summative focus was required to address their esteem needs.

The batch of 2006-08 has participated in the SET development. New students did not show much enthusiasm about SET. They did not understand the importance of feedback, and the Institute had no formal system to educate them in this regard. Besides, it was not mandatory for a student to give feedback. A focus group discussion with them revealed that they were not particularly happy about the end term SET as that hardly improved their ongoing courses in any way. Finally, the administration decided to address the concerns of the students and the teachers.

It is proposed to educate students about the SET before they evaluate the teachers. In order to discriminate between excellent and good teachers, the computation process for the overall faculty score would change. The faculty score would represent the sum of the key area scores instead of the average. The maximum possible score would be 60. The faculty score would be graded as per a grading scale depicted in Table 3. Assigning different weights to different performance dimensions is also on the agenda. These changes are likely to fulfill the esteem needs of the teachers to a certain extent. Midterm feedback consultations are also being contemplated to facilitate improvement of ongoing courses.

Conclusion

This article has presented the rationale and the processes concerning institution-specific SET for a very small management institution. Because it is institution-specific, it cannot be substantively compared with SET instruments of other institutions. It must be noted that these processes were situated in a relatively small and young organization where personal contact and informal interactions could largely substitute for the formal organization in many respects. These processes are expected to be much more complicated for large and established universities. The relative effectiveness of SETs based on behaviors, outcomes, traits, or judgments on broadly mentioned issues needs further study.

Involvement of students and faculty in the development process may be important for the success of SET. Post-implementation experiences with the instrument highlight the importance of linking SET with professional development opportunities. Simultaneously it is also apparent that a wholly formative SET might not contribute toward fulfilling the esteem needs of teachers. A limited summative focus appears to be justified. The experiences highlight the importance of open communication and a climate of trust for a successful SET. All new students need to be educated about SET and its impact on the quality of their academic life. It is more likely to lose effectiveness with time if it fails to reflect the changing needs of students, faculty, and administration. SET should be allowed to evolve rather than settle in certain grooves for a long time.

Table 3
Proposed Grading Scale for Overall Faculty Score

Faculty Grade (Acceptable)				Faculty Grade (Unacceptable)			
A+	A	B+	B	C+	C	D+	D
52.50-60.00	45.00-52.49	37.50-44.99	30.00-37.49	22.50-29.99	15.00-22.49	7.50-14.99	< 7.50

Endnote

CSREM is now a part of the School of Management of Centurion University of Technology and Management (CUTM), Odisha.

References

Arreola, R. A. (2007). *Developing a comprehensive faculty evaluation system: A guide to designing, building, and operating large-scale faculty evaluation systems (3rd Ed.).* Bolton, MA: Anker.

Aubrecht, J. D. (1984). Better faculty evaluation systems. In P. Seldin (Ed.), *Changing practices in faculty evaluation: A critical assessment and recommendations for improvement*, (pp. 85-91). San Francisco, CA: Jossey-Bass.

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education, 17* (1), 48-62.

Bisoux, T. (2007, May/June). The MBA reconsidered. *BizEd*. Retrieved from http://www.aacsb.edu/ publications/archives/mayjun07/44-49_bized.pdf.

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

Braxton, J., & Hargens, L. (1996). Variation among academic disciplines: Analytical frameworks and research. In J. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 11). New York: Agathon Press.

Cashin, W. E. (1988). Student ratings of teaching: A Summary of the Research. *Idea paper no. 20*. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.

Cashin, W. E. (1990). Students do rate different academic fields differently. In M. Theall and J. Franklin (Eds.), *Students ratings of instruction: Issues for improving practice: New directions for teaching and learning* (Vol. 43, pp. 113-121). San Francisco, CA: Jossey-Bass.

Cashin, W. E. (1995). Student ratings of teaching: The research revisited. *Idea paper no. 32*. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.

Cashin, W. E. (1999). Student ratings of teaching: Uses and misuses. In P. Seldin & Associates (Eds.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 25–44). Bolton, MA: Anker.

Centra, J. A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness.* San Francisco, CA: Jossey-Bass.

Clegg, S., & Ross-Smith, A. (2003). Revising the boundaries: Management education and learning in a postpositivist world. *Academy of Management Learning and Education, 2*(1), 85-98.

Comm, C. L. & Mathaisel, D. (1998). Evaluating teaching effectiveness in America's business schools: Implications for service marketers. *Journal of Professional Services Marketing, 16*(2), 163-170.

Cook, S. (1989). Improving the quality of student ratings of instruction: A look at two strategies. *Research in Higher Education, 30*(1), 31-45.

Dunegan, K. J. & Hrivnak, M. W. (2003). Characteristics of mindless teaching evaluations and the moderating effects of image compatibility. *Journal of Management Education, 27*(3), 280–303.

Eley, M., & Stecher, E. (1997). A comparison of two response scale formats used in teaching evaluation questionnaires. *Assessment and Evaluation in Higher Education, 22*(1), 65-70.

Ellis, L. Burke, D., Lomire, P. & McCormack, D. (2003). Student grades and average ratings of instructional quality. *Journal of Educational Research, 97*(1), 35-40.

Emery, C. R., Kramer, T. R. & Tian, R. G. (2003). Return to academic standards: A critique of students' evaluations of teaching effectiveness. *Quality Assurance in Education: An International Perspective, 11*(1), 37–47.

Feldman, K. A. (1988). Effective college teaching from the students' and faculty's view: Matched or mismatched priorities? *Research in Higher Education, 28*(4), 291-344.

Fernandez, J., Mateo, M. A. & Muniz, J. (1998). Is there a relationship between class size and student ratings of teaching quality? *Educational and Psychological Measurement, 58*(4), 596-604.

Fink, D. (2008). Evaluating teaching: A new approach to an old problem. In S. Chadwick-Blossey and D. R. Robertson (Eds.), *To improve the academy:*

Resources for faculty, instructional, and organizational development, (Vol. 26, pp. 3-21). San Francisco, CA: Jossey-Bass.

Franklin, J., & Theall, M. (1989). *Who reads ratings: Knowledge, attitudes, and practices of users of student ratings of instruction*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Geis, G. L. (1984). The context of evaluation. In P. Seldin (Ed.), *Changing practices in faculty evaluation: A critical assessment and recommendations for improvement* (pp. 101-107). San Francisco, CA: Jossey-Bass.

Gosling, J., & Mintzberg, H. (2006). Management education as if both matter. *Management Learning, 37*(4), 419-428.

Harrington, C., & Schibik, T. (2003). *Student evaluation of teaching: What every institutional researcher should know*. Paper presented at the 17th Annual Meeting of the Indiana Association for Institutional Research, Nashville, IN.

HBS Centennial Colloquia Report (2009). Business Summit: Business education in 21st century. Retrieved from http://hbswk.hbs.edu/item/6220.html.

Hutchings, P., & Shulman, L. (1999). The scholarship of teaching: New elaborations, new developments. *Change, 31*(5), 11–15.

Kim, C., Damewood, E. & Hodge, N. (2000). Professor attitude: Its effect on teaching evaluations. *Journal of Management Education, 24*(4), 458-473.

Knowles, L., & Hensher, D. A. (2005). The postgraduate business curriculum: The frontline in the war between professionalism and academic irrelevance. *International Journal of Management Education, 4*(3), 31-39.

Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response order effects in survey measurement. *Public Opinion Quarterly, 51(2)*, 201-219.

Krosnick, J. A., & Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. In L. Lyberg, M. Collins, L. Decker, E. Deleeuw, C. Dippo, N. Schwarz & D. Trewing (Eds.), *Survey measurement and process quality* (pp. 141-164). New York, NY: Wiley-Interscience.

Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, W. E., Kopp, R. J., et al. (2002). The impact of 'no opinion' response options on data quality. *Public Opinion Quarterly, 66*(3), 371–403.

Langbein, L. (2005). *Management by results: Student evaluation of faculty teaching and the mismeasurement of performance*. Paper presented at the Annual Meeting of Public Choice Society. Retrieved from http://www.pubchoicesoc.org/papers2005/langbein.pdf.

Lesser, D., & Ferrand, J. (2000). Effect of class size, grades given, and academic field on student opinion of instruction. *Community College Journal of Research and Practice, 24*(4), 269-277.

L'Hommedieu, R., Menges, R., & Brinko, K. (1990). Methodological explanations for the modest effects of feedback from student ratings. *Journal of Educational Psychology, 82*(2), 232-240.

Marsh, H. W., & Dunkin, M. (1992). Student's evaluation of university teaching: A multidimensional perspective. In J. C. Smart (Ed.) *Higher education: Handbook of theory and research (vol. 8, 143-233)*. New York, NY: Agathon.

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist, 52*(11), 1187-1197.

Marsh, H. W. (1982). The use of path analysis to estimate teacher and course effects in student rating's of instructional effectiveness. *Applied Psychological Measurement, 6*(1), 47-59.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76*(5), 707-754.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11*(3), 253-388.

Marsh. H. W. (1993). Multidimensional students' evaluations of teaching effectiveness. *Journal of Higher Education, 64*(1), 1-18.

Mateo, M. A., & Fernandez, J. (1996). Incidence of class size on the evaluation of university teaching quality. *Educational and Psychological Measurement, 56*(5), 771-778.

McKone, K. (1999). Analysis of student feedback improves instructor effectiveness. *Journal of Management Education, 23*(4), 396–415.

Mello, J. A. (2004). *Strategic human resource management.* Singapore: Thomson South-Western.

Miller, R. I. (1987). *Evaluating faculty for promotion and tenure.* San Francisco, CA: Jossey-Bass.

Murphy, T., Maclaren, I. & Flynn, S. (2009). Toward a summative system for the assessment of teaching quality in higher education. *International Journal of Teaching and Learning in Higher Education, 20*(2), 226-236.

Murray, H. G. (2005). *Student evaluation of teaching: Has it made a difference*? Paper presented at the Annual Meeting of the Society for Teaching and Learning in Higher Education. Retrieved from

http://www.mcmaster.ca/stlhe/documents/Student% 20Evaluation%20of%20Teaching.pdf.

Murray, H. G., Rushton, J. P., & Paunonen, S. V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology, 82*(2), 250-261.

Nasser, F., & Fresko, B. (2002). Faculty views of student evaluation of college teaching. *Assessment & Evaluation in Higher Education, 27*(2), 187–198.

Richardson, J. T. E. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education, 30*(4), 387-415.

Seldin, P., & Angelo, T. A. (1997). *Assessing and evaluating faculty: When will we ever learn? (To use what we know).* Proceedings of the American Association for Higher Education Conference on Assessment and quality assessing impact: Evidence and action.

Seldin, P. (1984). *Changing practices in faculty evaluation: A critical assessment and recommendations for improvement.* San Francisco: Jossey-Bass.

Seldin, P. (Ed.). (2006). *Evaluating faculty performance.* Bolton, MA: Anker.

Serva, M., & Fuller, M. (2004). Aligning what we do and what we measure in business schools: Incorporating active learning and effective media use in the assessment of instruction. *Journal of Management Education, 28*(1), 19-38.

Shevlin, M., Banyard, P., Davies, M. & Griffiths, M. (2000). The validity of student evaluation of teaching in higher education: Love me, love my lectures? *Assessment & Evaluation in Higher Education, 25*(4), 397-405.

Stapleton, R. J., & Murkison, G. (2001). Optimizing the fairness of student evaluations: A study of correlations between instructor excellence, study production, learning production, and expected grades. *Journal of Management Education, 25*(3), 269-291.

Theall, M., & Franklin, J. (Eds.) (1990). *Student ratings of instruction: Issues for improving practice. New Directions for Teaching and Learning (No. 43)*. San Francisco, CA: Jossey-Bass.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response.* New York: Cambridge University Press.

Waller, S. C. (2004). *Conflict in higher education faculty evaluation: An organizational perspective* [Electronic Version]. from http://www.newfoundations.com/OrgTheory/Waller721.html.

Wright, R. (2006). Student evaluations of faculty: Concerns raised in the literature and possible solutions. *College Student Journal, 40*(2), 417-422.

Younes, B. (2003). Faculty evaluation: Towards a happy balance between competing values. *World Transactions on Engineering and Technology Education, 2*(1), 117-120.

Zell, D. (2001). The market-driven business school: Has the pendulum swung too far?. *Journal of Management Inquiry, 10(*4), 324-338.

KUMAR ALOK is Assistant Professor of Organizational Behavior and Human Resource Management at Chandragupt Institute of Management Patna. He was associated for about three and half years with CSREM. He has several national and international journal and conference publications to his credit. His areas of research interest comprise education, leadership, organization theory, and Indian philosophy.

Appendix
Faculty Evaluation Sheet

Thoughtful student feedback can help improve teaching effectiveness. This tool is designed for that purpose and your cooperation is highly appreciated in this regard.

General Instructions

- Given below are 18 performance examples of your teacher for your rating.
- Please indicate your rating on each performance example by darkening  the appropriate circle.

FD	LD	SD	SA	LA	FA
Fully Disagree	Largely Disagree	Slightly Disagree	Slightly Agree	Largely Agree	Fully Agree

PE No.	Performance Example	Rating Scale					
		FD	LD	SD	SA	LA	FA
1	Gave assignments that were helpful in understanding the subject better.	○	○	○	○	○	○
2	Strictly adhered to the deadlines of assignment submission.	○	○	○	○	○	○
3	Encouraged students to think and to question.	○	○	○	○	○	○
4	Used to create a threatening environment in the class.	○	○	○	○	○	○
5	Asked students for suggestions regarding the course outline.	○	○	○	○	○	○
6	Used to look confused while teaching complex topics.	○	○	○	○	○	○
7	Used to answer students' questions clearly.	○	○	○	○	○	○
8	Used to briefly summarize the previous lecture at the beginning of each class.	○	○	○	○	○	○
9	Often said, "I have explained the topic. It is your problem if you have not understood it."	○	○	○	○	○	○
10	Used to take very interactive sessions.	○	○	○	○	○	○
11	Provided course outline having helpful suggestions regarding recommended books/websites, group formation, projects, evaluation pattern and general rules for the course.	○	○	○	○	○	○
12	Never made any attempt to make the class interesting.	○	○	○	○	○	○
13	Described the concepts and processes related to the topic with the fundamental logic behind them.	○	○	○	○	○	○
14	Offered to explain questions and their answers once exams were over.	○	○	○	○	○	○
15	Used to mention areas of improvement and the ways to improve while giving feedback to students.	○	○	○	○	○	○
16	Clearly explained the evaluation criteria to students.	○	○	○	○	○	○
17	Encouraged students to seek his or her help whenever in need.	○	○	○	○	○	○
18	Emphasized only the theoretical aspect of the subject.	○	○	○	○	○	○