

Assuring Teaching Quality Through Student Evaluation

by Asst. Prof. Dr. Kasma Suwanarak

Introduction

Improving teaching quality is not the only motive that every educational institute may have. Rather, many writers on teacher evaluation (e.g. Duke, 1995; Ingvarson, 2001; McLaughlin and Pfeiffer, 1988; Shinkfield and Stufflebeam, 1995; Stronge, 1995) concur that the basic purpose of teacher evaluation is to guarantee teaching quality or ensure that students are well taught. As Coble and Piscatelli (2002) and Goldrick (2002) note, teaching quality has the greatest impact on students' achievement. Therefore, educational institutions assure teaching quality by evaluating basic teaching competence and ensuring that the right people are promoted to better positions (Gibbs, 1995). Taking this into account, Shinkfield and Stufflebeam (1995) critique that this can create tension in a workplace; teachers whose teaching quality is not of a high standard will be under pressure as this might impact on their professional development. In order to achieve the purposes of teacher evaluation, the existing mechanisms of quality assurance need to be improved and developed in order to create meaningful evaluation of teaching. This will contribute to improvement in various aspects of teaching, extensive implementation of professional development, and consequently help raise student achievement.

Evaluation

In educational research, evaluation is considered as a method of acquiring information and processing the evidence needed to determine the student's level of learning and the effectiveness of the teaching (Daresh and Playko, 1995; Wanzare, 2002). However, many educators have a similar view that evaluation entails more interesting issues. According to Nunan and Lamb (1996), evaluation involves deciding what aspects in relation to teaching practice should be changed. It is a system of feedback that gives some of the information necessary to determine where improvement is needed. Likewise, Rea-Dickens and Germaine (1998) view evaluation as an intrinsic part of teaching and learning. From this general perspective, evaluation provides a wealth of information for the future direction of classroom practice. Nisbet (1987), in agreement with this, adds that evaluation makes a substantial contribution to the understanding and proper consideration of education change and development.

Students as sources of evaluation of teaching quality

Basically, students' views on teachers' performance and learning outcomes are considered the major source for evaluating teaching. Ellington (1993) underlines that a critical study of learning outcomes can be of great assistance in the process of teacher evaluation when an instructional system or a course has clearly defined objectives or learning outcomes. According to Tucker and Stronge (2000), it is also important to note that the patterns of gain or decline in student achievement over time can be very useful for teacher evaluation as such patterns increase reliability from a statistical viewpoint and credibility from the teacher's perspective. Having had similar views regarding the significance of student achievement data, however, Allen (1999) gave a note of caution in using such data because of the variety of other factors that may affect student performance in tests.

Students play an important role in teacher evaluation since they are the most directly affected by the quality of teaching (Marsh, 1984). They are the only individuals who regularly observe teachers in action. This view is supported by Wilkerson et al. (2000), who note that obtaining feedback from students regarding their experiences with and their opinions of an instructional system is one of the most common approaches to evaluation. On account of this, it is very likely that students are in a position to discriminate between teacher performance and their own learning as well as they are able to evaluate specific aspects of classroom teaching (Arreola, 1995). However, Peterson et al. (2000) make a claim that students cannot respond to things they have no experience in. Brinko (1993), in agreement with this point, adds that students' feedback should be combined with advice from a faculty development specialist in order to help teachers improve teaching effectively.

Student rating questionnaires

Student evaluation of teaching is an integral part of higher education practice; student ratings of teaching effectiveness are the most commonly used method (Braskamp and Ory, 1994; Cohen, 1981; Feldman, 1988; Marsh, 1987) and are widely validated by students, faculty, and administrators (Centra, 1979). In the UK, student feedback questionnaires are widely used for the purpose of quality assurance of modules and programmes. Quantitative data obtained is also used for formative purposes: as supporting evidence in external teaching quality assessment, in cases made by individuals for promotion on the grounds of teaching excellence, and recently in portfolios of evidence of teaching competence for teacher accreditation (Coffey and Gibbs, 2001; Milman, 2002). Information from student evaluation, therefore, serves as a guide for potential changes in course materials and delivery methods (Shelvin et al., 2000).

In the USA, information from student evaluations of teaching can be used for decisions about conditions of employment such as salary and promotion (Shelvin et al., *ibid.*). In Canada, for instance, 94% of universities use student rating questionnaires alone or in conjunction with some other methods to evaluate teaching (Saroyan and Amundsen, 2001). In Thailand, the evaluation of teaching is a mixture of the practice employed in the UK and the USA. The information obtained from student ratings is useful for the improvement and development of learning and teaching practice, and the decision making of academic administrators about promotion and salary increase (The Ministry of Education, 2004).

Factors affecting student ratings

Several extraneous factors may influence the results of student rating questionnaires (Greenwald and Gillmore, 1997). The attractiveness of the teacher frequently influences student ratings. Accordingly, the term '*halo effect*' (as cited in Watkins and Durvasula, 2003) is used to describe the tendency of raters to attribute positive characteristics to attractive people. This effect can also be negative; an unfavorable general impression will lead to low ratings, even in areas where performance is strong. Students may give better ratings to more attractive teachers (Watkins and Durvasula, *ibid.*). Because of this effect, student ratings make less differentiation between '*strengths*' and '*weaknesses*' than is desirable.

Another interesting issue is proposed regarding teachers' leniency. Most researchers (Abrami et al., 1990; Arreola, 1995; d'Appolonia and Abrami, 1997; Greenwald and Gillmore, 1997) agree that teachers' leniency in grading may produce some prejudice in student ratings. Different abilities of students to rate effectively indicate another interesting factor affecting student ratings. Marsh and Dunkin (1992) assert that grades or student achievement and ratings have a small, but consistent and positive relationship. Students who learn more will receive higher grades and will be motivated to give higher ratings. However, Newport (1996) argues that students who obtain higher grades are unnecessarily more qualified raters than those who receive lower grades and that senior-level students are not essentially more qualified raters than university freshmen (*ibid.*).

Husbands (1998) remarks that reasons for taking course and previous attitude, including interest in a particular course, do in fact correlate slightly with student ratings of teaching. Interestingly, alumni still give the same ratings of teaching as they did when being students (McKeachie, 1996). Class size can be another factor which relates to the effectiveness of student ratings. From the study of Centra (1979), very small classes, those fewer than 15 students, are most highly rated while classes with 50-100 students receive the lowest ratings.

Reliability and validity of student ratings

According to Marsh and Roche (1997), student evaluations are reliable if they yield consistent patterns and stable results over time. There are three types of reliability which relate to student ratings. First, inter-rater reliability refers to agreement of all student ratings of one instructor. It provides the most common and appropriate indication of the reliability of student rating forms. Second, test-retest reliability refers to whether or not an instructor receives similar ratings every semester. Third is generalisability which reflects teaching effectiveness in general, not just in a particular course.

Validity is the degree to which a test actually measures what it is supposed to measure. The validity of student ratings indicates to what extent the rating items measure certain aspects of teaching effectiveness (Scriven, 1993). On account of this, Marsh (1984) claims that student evaluation is the only indicator of teaching effectiveness because the validity of student ratings has been thoroughly and rigorously established. There are three basic types of validity inherent in student ratings: content validity, construct validity, and criterion validity. Content validity incorporates estimates of the extent to which the content of rating scales relate to what it is designed to measure. Regarding construct validity, student ratings should correspond with the specific scales chosen to represent teaching effectiveness. Criterion validity represents performance in relation to particular task objectives.

Interestingly, a different viewpoint about student ratings of teaching is proposed by Darling-Hammond et al. (1983). While having an agreement on a high degree of reliability of student ratings, they remark that questions of validity and bias remain. Their point is supported by many studies, such as those of Abrami et al. (1990), Cohen (1983), Marsh et al. (1979), which indicate that student evaluations are a valid means of assessing student satisfaction. Dowell and Neal (1983), on the other hand, remark that student ratings are inaccurate indicators of student learning; they can only be best regarded as indices of '*consumer satisfaction*' rather than teaching effectiveness.

Therefore, according to Nuhfer (2003), student ratings should not be the only criteria for rating teaching effectiveness. Another interesting objection is that student ratings cannot be considered valid because students are incapable of evaluating good teaching; student evaluation represents nothing more than a popularity contest (The Center for Teaching and Learning, 1994). Despite criticisms regarding validity and bias, to some extent, student rat-

ings are reliable and valid to justify the effectiveness of teaching and to be one of several sources of information for administrative purposes.

Achieving efficiency in the evaluation of teaching

It is extremely important that explicit criteria for evaluating teaching be established and communicated to everyone involved in this practice. In agreement with Rea-Dickins and Germaine (1998), specific criteria help focus on particular aspects of teaching and lessen bias, with the result that quality of objectivity and constructive ideas for improving teaching and developing courses will be obtained. This taken into account, it is possible that the dean, department chairs, and academic administrators play a role in formulating the evaluation criteria in matters of systematic evaluation of teaching to guarantee quality of teaching practice. Additionally, the relative weight and a broad range of data sources may be explained and periodically discussed among teacher colleagues. Once the evaluation is well performed under the standardised criteria, it will help strengthen the system of quality assurance of teaching excellence.

It is most important to carefully construct an evaluation tool for the benefit of teaching and course improvements. Also, a current evaluation form has to be examined and improved regularly in order to ensure that it can profitably gather the most useful information according to the purpose of evaluation. Moreover, a great change for the better in evaluation procedure could be achieved if the evaluation of teaching is carried out comprehensively and systematically. In the Thai university context, the evaluation is usually carried out at the end of courses. The evaluation of teaching is only used for summative purpose as it is likely to focus on judgements about the effectiveness of teaching, but less so for the formative purpose in terms of constantly improving teaching. If the evaluation of teaching were also carried out mid-semester, teachers could consider early feedback as constructive criticism and use it for making improvements during the current semester (Ballantyne et al., 2000; Cuseo, 2000). In this way, teachers can see whether their teaching methods best contribute to students' understanding and whether there is any improvement in teaching strategies by comparing with summative feedback at the end of semester.

With regard to the procedure of collecting feedback, students should be reminded that evaluating teaching is an opportunity for them to provide meaningful information that could improve teaching quality and courses for them and future students. As specific comments and suggestions are considered valuable for improving teaching quality and course weaknesses, students should be encouraged and instructed how to give lengthy and useful responses to open questions. With a guarantee of anonymity, students will be more confident to provide the real feedback on teachers' teaching performance (Cuseo, 2000).

It is a fact that subjectivity regarding students' feedback on teaching is inevitable; however, it can be controlled and reduced by having more careful consideration of data over a range of teaching performance and over a period of time. In appropriate ways, effective evaluation of teaching should be broadened by including multiple sources of data (e.g. class observation by teacher colleagues and academic administrators, teachers' self-evaluation, interviews with current students and alumni) as supplements to students' rating feedback and written comments. Once feedback on teaching performance is derived from various sources, accurate data of teaching performance and a more acceptable level of reliability can be achieved as well as a fairer and more reasonable evaluation system can be developed (Arreola, 1995).

References

- Abrami, P. C., d'Apollonia, S. and Cohen, P. A. (1990). Validity of student ratings of instruction: what we know and what we do not. *Journal of Educational Psychology*, 82, 219-31.
- Allen, M. (1999). *Student Results and Teacher Accountability*. ECS Policy Brief.
- Arreola, R. (1995). *Developing a Comprehensive Faculty Evaluation System*. Bolton, MA: Anker Publishing Company.
- Ballantyne, R., Borthwick, J. and Packer, J. (2000). Beyond student evaluation of teaching: identifying and addressing academic staff development needs. *Assessment and Evaluation in Higher Education*, 25(3), 221-236.
- Braskamp, L. A., Brandenburg, D. C. and Ory, J. C. (1984). *Evaluating Teaching Effectiveness: A practical guide*. London: Sage Publications.
- Brinko, K. T. (1993). The practice of giving feedback to improve teaching: What is effective? *Journal of Higher Education*, 64, 574-593.
- Centra, J. A. (1979). *Determining Faculty Effectiveness*. London: Jossey-Bass.
- Coble, C. R. and Piscatelli, J. (2002). Teaching quality: a national perspective. *State Education Leader*, 20(3).

- Coffey, M. and Gibbs, G. (2001). The evaluation of the student evaluation of educational quality questionnaire (SEEQ) in UK higher education. *Assessment and Evaluation in Higher Education*, 26(1), 89-93.
- Cohen, P. A. (1981). Student rating of instruction and student achievement: a meta-analysis of multisection validity studies. *Review of Educational Research*, 51, 281-309.
- Cohen, P. A. (1983). Comment on 'a selective review of validity of student ratings of teaching'. *Journal of Higher Education*, 54, 448-458.
- Cuseo, J. (2000). *Evaluating New Student Seminars and Other First-Year Courses via Course Evaluation Surveys: Research-based recommendations regarding instrument construction, administration, data analysis, data summary and reporting results*. National Resource Center. Retrieved August 16, 2011 from http://www.sc.edu/fye/resources/assessment/evaluation/joe_cuseo.html
- d'Appolonia, S. and Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist*, 52(11), 1198-1208.
- Dares, J. and Playko, M. (1995). *Supervision as a Proactive Process, Concepts, and Cases* (2nd Edition). Prospect Heights, IL: Waveland.
- Darling-Hammond, L., Wise, A. E. and Pease, S. R. (1983). Teacher evaluation in the organisational context: a review of the literature. *Review of Educational Research*, 53(3), 285-328.
- Dowell, D. A. and Neal, J. A. (1983). The validity and accuracy of student ratings of instruction: a reply to Peter A. Cohen. *Journal of Higher Education*, 54, 459-463.
- Duke, D. L. (1995). The move to reform teacher evaluation. In D. L. Duke (Ed.), *Teacher Evaluation Policy: From accountability to professional development* (pp. 1-12). Albany, NY: State University of New York Press.
- Ellington, H. (1993). Evaluating the effectiveness of the teaching/learning process. In F. Percival, H. Ellington and P. Race (Eds.), *Handbook of Educational Technology* (3rd Edition). London: Kogan Page.
- Feldman, K. A. (1988). Effective college teaching from students' and faculty's view: matched or mismatched priorities? *Research in Higher Education*, 28, 291-344.
- Gibbs, G. (1995). Promoting excellent teaching is harder than you'd think. *Change*, 27(3), 16-20.

- Goldrick, L. (2002). Improving teacher evaluation to improve teaching quality. *Issue Brief*. NGA Center for BEST PRACTICES. Retrieved August 20, 2011 from <http://www.nga.org/cda/files/1202IMPROVINGTEACHEVAL.pdf>
- Greenwald, A. G. and Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52(11), 1209-1217.
- Husbands, C. T. (1998). Implications for the assessment of the teaching competence of staff in higher education of some correlated of students' evaluations of different teaching styles. *Assessment and Evaluation in Higher Education*, 23(2), 117-140.
- Ingvarson, L. C. (2001). *Strengthening the Profession: A comparison of recent reforms in the USA and the UK*. Australian College of Education Seminar Series, ACE: Canberra.
- Marsh, H. W. (1984). Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76, 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. and Dunkin, M. J. (1992). Students' evaluations of university teaching: a multi-dimensional perspective. In J. C. Smart (Ed.), *Higher Education: Handbook of theory and research* (pp. 143-233). New York: Agathon Press.
- Marsh, H. W., Overall, J. U. and Kesler, S. P. (1979). Validity of student evaluations of instructional effectiveness: a comparison of faculty self-evaluations and evaluation by their student. *Journal of Educational Psychology*, 71, 149-160.
- Marsh, H. W. and Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist*, 52(11), 1187-1197.
- McKeachie, W. J. (1996). Student ratings of teaching. *American Council of Learned Societies Occasional Paper*, No. 33. Retrieved September 9, 2011 from <http://www.acls.org/op33.htm#McKeachie>
- McLaughlin, M. and Pfeiffer, R. S. (1988). *Teacher Evaluation, Improvement, Accountability, and Effective Learning*. New York Teacher College Press.
- Milman, N. B. (2002). *Web-Based Digital Teaching Portfolios: What happens after they graduated?* Paper presented at the annual meeting of the American Educational Research Association, April 2002, New Orleans, LA. Session#7.09. Retrieved August 24, 2011 from http://home.gwu.edu/~nmilman/milman_AERA_2002.pdf

- Newport, J. F. (1996). Rating teaching in the USA: probing the qualifications of student raters and novice teachers. *Assessment and Evaluation in Higher Education*, 21(1), 17-22.
- Nisbet, J. (1987). The role of evaluation in accountability systems. In R. Murphy and H. Torrance (Eds.), *Evaluating Education: Issues and methods*. London: Harper and Row.
- Nuhfer, E. B. (2003). *Of What Values are Student Evaluations?* Center for Teaching and Learning. Idaho State University. Retrieved September 2, 2011 from <http://www.isu.edu/ctl/facultydev/extras/student-evals.html>
- Nunan, D. and Lamb, C. (1996). *The Self-Directed Teacher*. Cambridge: Cambridge University Press.
- Peterson, K., Wahlquist, C. and Bone, K. (2000). Student surveys for school teacher evaluation. *Journal of Personnel Evaluation in Education*, 14, 135-153.
- Rea-Dickins, P. R. and Germaine, K. P. (Eds). (1998). *Managing Evaluation and Innovation in Language Teaching: Building bridges*. London: Longman.
- Saroyan, A. and Amundsen, C. (2001). Evaluating university teaching: time to take stock. *Assessment and Evaluation in Higher Education*, 26(4), 341-353.
- Scriven, M. (1993). The validity of student ratings: in teacher evaluation. *Evaluation and Development Group*.
- Shevlin, M., Banyard, P., Davies, M. and Griffiths, M. (2000). The validity of student evaluation of teaching in higher education: love me, love my lectures? *Assessment and Evaluation in Higher Education*, 25(4), 397-405.
- Shinkfield, A. J. and Stufflebeam, D. L. (1995). *Teacher Evaluation: Guide to effective practice*. Boston Kluwer Academic Publishers.
- Stronge, J. H. (1995). Balancing individual and institutional goals in educational personnel evaluation: a conceptual framework. *Journal of Personnel Evaluation in Education*, 21, 131-151.
- The Center for Teaching and Learning. (1994). Student evaluation of teaching. *For Your Consideration: Suggestions and reflections on teaching and learning*, 16, September. University of North Carolina at Chapel Hill.

- The Ministry of Education. (2004). *Educational Standards and Quality Assurances: Education in Thailand 2004*. Retrieved May 23, 2011 from <http://www.edthai.com/Publication/edu2004//edu2004.pdf>
- Tucker, P. and Stronge, J. (2000). Achievement and accountability: How should student achievement test scores be used-fairly-in evaluating the performance of educators? *Virginia Journal of Education*, March.
- Wanzare, Z. O. (2002). Rethinking teacher evaluation in the third world: the case of Kenya. *Educational Management and Administration*, 30(2), 213-229.
- Watkins, R. and Durvasula, R. (2003). Students evaluating professors: style over substance? Retrieved August 20, 2011 from <http://www.studentgroups.ucla.edu/upj/fall2003/reynoldwatkins.html>
- Wilkerson, D., Manatt, R., Rogers, M. and Maughan, R. (2000). Validation of student, principal, and self-ratings in 360 degree feedback for teacher evaluations. *Journal of Personnel Evaluation in Education*, 14(2), 179-192.

Asst. Prof. Dr. Kasma Suwanarak holds a doctoral degree in TESOL (Teaching English to Speakers of Other Languages) from the University of Exeter, UK. She obtained her master's degree in Applied Linguistics from Kasetsart University, where she also completed her first degree in English Literature from the Faculty of Humanities. Presently, she is the Associate Dean of the Graduate School of Language and Communication, National Institute of Development Administration (NIDA).