The “Evaluation” of Teaching in a Japanese University Context and its Relation to Students’ Beliefs about “Effective” English Teachers

Peter Burden

Abstract

Student evaluation of classes has become established in Japanese universities despite lingering doubts among English Language educators about reliability, validity, and utility of the evaluation instrument. This study questions whether results of a closed-item questionnaire given to students who have had just one semester in a university can be considered suitable criteria for decisions which can affect teachers’ futures including the issue of retention of part-time teachers. In this paper, the controversial history of student evaluation and its introduction into Japan is reviewed, and the results of a quantitative survey administered to 307 students are discussed. The survey asks whether students agree that filling in evaluation forms is a worthwhile practice; what should happen to teachers who receive poor evaluation; what students think are the purposes of evaluation; and what students think they should evaluate. The paper also links notions of evaluation with student attributes of good teachers as students were also asked to list attributes of “good” teachers. While findings are necessarily limited being based on quantitative data, findings do suggest that students may evaluate in terms of teacher characteristics such as enthusiasm.
Implications are that everyone involved in evaluation—administrators, teachers, and students—need to consider issues of validity including how student ratings influence teaching and organizes the learning environment to avoid it becoming merely a purposeless ritual.

学生による授業評価が、教育者側がその信頼性、有効性、活用性に関して疑問視しているにもかかわらず、日本の大学で広く行なわれるようになった。学生は大学で1セメスターのみで授業を受けないので、その学生が記入した選択式の授業評価アンケートの結果が、はたして非常勤講師が仕事を保持できるかどうかなどの教師の将来に影響を与えるか否かという点を本研究は問っている。本稿では、学生の授業評価に関する今までの議論や日本での授業評価の導入を再検討し、307名の学生を対象に行なった量的調査の結果を論ずる。調査では、学生が授業評価の記入をする価値のあるものと考えているかどうか、評価が悪かった教師はどうなると思うか、評価の目的は何だと思うか、何を評価すればいいと思うかなどを尋ねた。また本稿では、学生が「よい教師像」を挙げるように質問を作り、評価概念と学生が抱く良い教師像を連絡している形をとっている。本研究の結果は、学生が熱勁なような教師の特徴という点で評価しているかもしれないことを示唆している。授業評価に関わっているすべての人、つまり学校側、教師、学生、この三者が、授業評価の有効性について熟考する必要があることを示唆しており、学生の評価がいかに教授法に影響を与えるものであるか、また学習環境が単なる目的のない慣例の場になるのを避けるためにいかにその環境を構成するのかを考えるべきとしている。

Introduction

How can we recognize effective teaching and good teachers? One way is to evaluate teachers by their ability to facilitate good academic work in their learners concentrating on the products or outcomes of teaching. This relies on incentives based on examination results. Another way is valuing ability to encourage personal change and development in their students (Shevlin, Banyard, Davies and Griffiths, 2000) and the encouragement of motivation towards life long learning. This requires time and is hard to quantify. A third way is through the use of student ratings, which has the advantage of being measurable and observable,
cheap to implement, and represents the processes that are believed to have a positive or negative impact on learning (Centra, 1979; Feldman, 1997). Many teachers would welcome student evaluation as one aspect for formative professional development to allow competent teachers to achieve higher levels of professional competence and expand their understanding of self, role, context, in the realization that the very nature of professional knowledge is based on systematic enquiry.

The Education Ministry (henceforth "MEXT") in Japan has made implementation of self-evaluation compulsory since 1999 (MEXT, 2004). However, in many Japanese universities, the only form of evaluation is end of semester summative evaluation which suggests minimal impact on teacher development. The recent flurry of activity is the result of a number of factors including the rise of consumerism as a form of market orientation as a result of criticism of university education and the public demand for institutional accountability (Arimoto, 1997). Universities are now encouraging the system of student-as-consumer evaluating teachers in the belief that popular teachers and courses offer student satisfaction and will attract potential students. Internationalization and declining population demands that university quality has competitive value and evaluation is seen as a cheap and quick way to quantify teaching quality. The changing nature of employment and the disappearance of lifelong employment means that students need marketable skills so necessary to examine the total quality of the education process. A fourth, highly contentious potential use is for program and personnel decisions where evaluation can be used as convenient evidence in an era of downsizing in university faculty leading up to the 2009 "universalization" of access.

Teachers resignedly administer ratings as part of the end of semester ritual alongside summative testing of students. The Education Ministry in Japan states that in 2001 100% of "National" universities, 92% of private universities and 82%
of public universities had implemented self monitoring and self evaluation which includes student evaluation (MEXT, 2004).

These evaluation forms usually ask students to rate specific teacher characteristics such as "teacher enthusiasm," "knowledge of the subject," and whether classes started and finished "on time" along with a final, global characteristic of "overall effectiveness" of the instructor. The ratings are usually through a 5-point Likert-type scale with the following anchors: "very poor (1)," and "very good (5)." As noted above, one issue among many is whether these overt variables are measured because they are important in effective teaching, or simply because they happen to be observable and therefore measurable. Arguably, current usage of student ratings encourages school administration to rate teachers based on characteristics of teaching not necessary to good teaching, but merely associated with it. Student ratings concentrate on overt, surface techniques visible to the students which often suggest that competent teachers need only good communication skills without acknowledging principles of cognitive development, motivation or pedagogical skills necessary to help students understand the curriculum (Good and Mulryan, 1990). The whole range of underlying processes that are not observable thus remain unexplored (Saroyan and Amundsen, 2001).

Criticisms of the use of end of semester ratings related to student reactions are widespread. It has been frequently demonstrates in the literature that a halo effect occurs when appraising teachers, Shevlin, Banyard, Davies, and Griffiths (2000) arguing that ratings:

Were demonstrated to be significantly effected by the students' perception of the lecturer on a variable that should be unrelated to assessments of teaching ability (p. 402).
Williams and Ceci (1997) point to content–free stylistic changes such as voice pitch and gestures having a dramatic change on ratings including the seemingly unrelated positive student approval of texts used in the class lifetime. It is suggested that ratings reflect the way students feel as well as the way they think, and so not only ratings for teacher enthusiasm go up but other effects as well (Kulik, 2001). Therefore, teachers with "low" ratings should concentrate on their greater relative weakness, which would then pull up the other variables as evaluations of effectiveness are closely linked to global notions of overall student satisfaction.

For student evaluation to be utilized by teachers for formative development, the data must be credible and that the participants involved can place trust in the evaluation process and the results of the evaluation (Braskamp and Ory, 1994). Getting credible data means attaching importance to gaining usable data and ensuring that teachers accept student judgements. Abrami, d’Appollonia, and Rosenfield (1997) point to the practical and theoretical utility of ratings depends on the extent to teachers believe students are uniform in their ratings of teachers, and are free from biasing influences. All too often ratings forms ask students questions they cannot possibly answer (see Seldin, 1993). We should not ask questions that require professional high inferential knowledge. For this reason, many advocate multiple data collections such as peer review to gain different perspectives.

The Introduction of student ratings into Japan

Kitamura (1997) noted that until the 1980s higher education enjoyed unprecedented growth and prosperity because of ever increasing enrollment resulting from two post-war baby booms and continued high employment demands for college graduates. As the Education Ministry adopted a laissez faire policy, universities were not well–organized communities but often only a
congregation of extremely individualistic professors. Usually, there was little communication among the chairs of faculties while the lectures were and left entirely to the professors, who could freely change the structure and contents of their lectures even in the middle of the courses. (Shimizu, Baba, & Shimada, 2000). However, seismic shifts in education policy geared towards accountability and market reform have reflected rapid societal change. There have calls for reform since at least the mid 1980s when *The Japan Association of Private Colleges and Universities* (1987) noted the "widespread unease and dissatisfaction with the present system... the all importance of scholastic aptitude test scores, [and] perfunctory lectures at universities" (p. 212) The Association suggested that a casual glance at any contemporary newspaper or magazine would "find reports on these and other topics indicative of society's discontent with education." Behind such a pronouncement of dissatisfaction is the widely-held view that the Japanese university is an "educational disaster—a four year "holiday" (Tsuda, 1993), or a "moratorium period" (Sugimoto, 2003, p. 139) where little of value or significance is actually learned or accomplished (Tsuda, 1993). High schools and corporate training were seen to provide the education necessary and "stringent schools" were contrasted with "slack universities" (Sugimoto, 2003, p. 139). The Japanese university was portrayed as "an enigmatic and glaring aberration in a generally excellent and effective system of education" (Tsuda, 1993, p. 305) while tolerated as a societal rite of passage. But as the twenty-first century has become a time of "intelligence restructuring" (Tsurata, 2003, p. 126); the economic downturn has lead to changes in demand for human resources. The economy and society seen to be borderless while the progression of "informatization" (sic) creates changes in abilities that employers are seeking in potential employees (see MEXT, 2004). Japan, as a post-industrial democratic society, is now in search of education for a new type of individual, fit for age of "globalization" (Saito & Imai, 2004). To compound matters, the eighteen year-old population has declined
by 20% since the peak of around 2 million in 1992 and is expected to fall by another 20% to 1.2 million by 2009. Estimates are that the total capacity of universities to accommodate new entrants will reach 100% in 2009. In 1995 46.2% of the 18 year old cohort were recorded as enrolled, and is expected to exceed 50% (Arimoto, 1997) in the near future. Reforms of teaching will have to cope with such educational massification resulting from the diversification of students, as almost all students who wish to enter university will. A new generation of eighteen years olds less competent in university studies will enter university, lowering academic standards as they lack fundamental knowledge and the intellectual curiosity to study so the number of establishments offering remedial seminars is expected to continue to rise. While learning ability and learning motivation had deteriorated in the past five years (Yamada, 2001), Saito and Imai (2004, p. 590) point to a “nihilism in learning... [whereby] many cannot find any hope in what they learn; they feel that nothing makes a difference in their lives [and] have not experienced the joy of liberation through learning even if freedom has increased in quantitative terms.” The 1999 curriculum revision at junior high level which introduced Integrated Studies has added to the concern over standards as the curriculum has effectively been reduced by 30%. Only a single page of the junior high curriculum is devoted to guidelines how to teach the new Integrated Studies potentially impacting on students’ ability to perform in English classes at universities. The topic based exploratory learning approach is seen to cut across traditional subject areas and would lead to a 25% decline in English if schools were teaching the maximum of 140 hours a year (Cave, 2001).

Against this backdrop the number of private institutions has continued to increase from 59.7% in 1955, to 73.2% in 1996 (Arimoto, 1997) and these institutions depend for a large part of income on tuition fees. This heavy dependence on fees means that many private institutions are facing prospect of bankruptcy in this time of retrenchment if they cannot attract sufficient students.
The *Asahi Shinbun* noted in July 2005 that 29% of four-year universities and 45% of two-year institutions were operating at a loss. Therefore, the student is both the customer and the consumer of educational services.

Accordingly, there is a growing need for the university to promote autonomy and distinctiveness and to improve the quality of education and research to attract students as choosers within a heavily skewed buyer's market where students are "courted customers" rather than "supplicants for admission" (Kitamura, 1997, p. 147). At the same time, an emerging reform movement requires that post secondary education become more economic centered, more market sensitive and therefore influenced by government shifts towards decentralization encouraging deregulation, freedom, flexibility, individualization and accountability based on business values and market principles. MEXT (1998) suggests that universities face strong criticism as education content and current reform methods do not reflect changes in the student population while the perceived fall in the quality of education has led to an inability to meet student learning wishes. Doyon (2001) adds that the highly bureaucratic educational system has been slow to respond to needs of the industrial sector producing obedient workforce that could be trained easily in the company. This does not reflect a quickly changing industrial situation facing the prospect of mid-career hiring and firing graduates with marketable skills. A University Council report entitled *A Vision for Universities in the 21st Century and Reform Measures* was unleashed (MEXT, 1998) which placed responsibility on educators to make ongoing efforts to develop classes that stimulate enthusiasm for learning in their students as a means of "improving" teaching. This view suggests that universities are failing to deliver the products the consumer demands such as easy to learn content, so the teachers need to be held more accountable through more rigid forms of appraisal. The emphasis in education on economic rationalization and economic accountability has as an attendant requirement that teachers yield to notions of standardization, efficiency
and effectiveness (see, for example, Smyth, 1988). These are clear indicators of an intensification of MEXT's ideological control in contrast to the laissez faire policy noted earlier with shifts in the previously egalitarian policy of distributing subsidies among institutions on an equal basis to one of concentrating subsidies on institutions that are perceived to offer high quality education (Amano, 1997).

Control over teachers has been further tightened through a contract system for personnel of five years, introduced to remove "habitual practices of life-long employment, seniority rule, [and] academic nepotism (MEXT, 2004). It is claimed that "increase[d] mobility" of educators will "revitalize" education through fostering a competitive environment by third party evaluation (MEXT, 2003). However, this can only further reduce teacher morale, already weakened by the increased demands from the popularization of education.

This popularization has been reflected in the "seven priority stages" for education outlined in the Educational Reform Plan for the 21st Century (MEXT, 2001) whereby teachers must offer "easy to understand classes" that are "enjoyable." However, the report has not laid out any conceptions of teaching that underpin such requirements. Similarly, while "appropriate measures" are to be taken against teachers who are seen to "lack teaching abilities" including suspension, there is a lack of information on what constitutes "ability," especially from an English teaching perspective, nor any diagnostic advice on how to "improve" this perceived inability.

Thus an economic centered, more market sensitive government shift towards decentralization in Japan coupled with a belief that good teaching "can be defined in terms of generic skills or actions" (Pratt, 1997, p. 25) suggests a worrying situation within tertiary EFL (English as a Foreign Language) vis-à-vis teacher accountability to the student "stakeholders" in education and the demands of an "audit culture" (Poole, 2005). Through a market agenda for universities performance indicators and quality assurance are paramount (MEXT, 1998), so a
great emphasis is placed on student evaluation. This is despite teacher skepticism over the imposition of evaluation from the university hierarchy who are unfamiliar with teaching practices in EFL education and who believe teaching is carried out through a lecture style. This leads to little sense of teacher ownership or involvement in the process. The punitive nature of evaluation means that minimum standards are encouraged, while good or excellent results are seldom commented on or taken into account when teachers seek career advancement (Harvey, 2003). As these summative evaluations are usually used in isolation, students become the sole determinant of a teacher’s “success” or “failure” with implications for retention or dismissal.

Many universities write their own evaluations which are administered across the subject range regardless of subject content, teaching approach, methods, and class size. As subject area teachers have their own conceptions of teaching which vary when fronting large lectures of three hundred students or during personalized seminars of just a few students, this means that comparatively evaluating teachers according to some externally imposed norm is both complex and of dubious value. There is a need to consider all the different learning tasks that different subject-specialists use. If evaluations are to be rigorous and credible they must acknowledge the essential and substantive aspects of EFL teaching, rather than just the most common attributes of teaching and learning (Pratt, 1997). How can students judge overall effectiveness of teachers when so many variables can influence this effectiveness, and while there is still no agreement on what constitute notions of effective teaching?

**Previous studies of students’ views on the evaluation process**

Implicit in evaluations is the assumption that students fill in anonymous instruments honestly and importantly, willingly. As student responses are the sole determinants of faculty member’s evaluation, to what degree is there an
assumption that students carefully read the questions, have reflected back over the course of the whole semester and made rational, thoughtful decisions? Many assumptions are implicit when students are “asked” for their input, but do students share the sense of responsibility?

There have been few studies asking students how conscientiously they respond to the questions, whether they take the whole process seriously, and what purposes they think are served by the evaluations. The lack of a transparent purpose of evaluation relayed to students as to whether evaluations are formative for development or summative for judging worth is crucial as Spencer and Schmelkin (2002, p. 439) note, “since students are unsure whether their opinions matter, or to what purpose the ratings are put, they may not pay attention to them in contrast to stated desire to provide feedback.” Are students mindful to complete evaluations in a meaningful way or are they “simply discharging a boring chore”? (McKeachie, 1997a, p. 1223). Dunegan and Hrivnak (2003) talk of “mindless” evaluation which leads to a situation where:

Information processing is limited, attention to detail is reduced, and our cognitive representation of the context is simplified. When mindless, we respond like automatons, drawing on scripts and schema in the past to define the present (p. 282).

Dunegan and Hrivnak (2003) point to three factors that may influence mindlessness in evaluation. Firstly, evaluation is cognitively taxing and emotionally demanding yet students at the end of term when evaluations are typically administered is a time when students are “inundated with other responsibilities” (p. 283) such as class tests or presentations that also demand cognitive attention and hence cognitive distraction on evaluations. Secondly, expectancy and equity theories of motivation suggest that students will only invest effort believing it will produce additional benefits but as evaluation completed
anonymously and because students are told their grades will not be affected there is little incentive to diligently fill in the forms. Thirdly, if evaluation becomes routine, and if it is customary for students to complete evaluations for many classes the process has become so standardized that "mindlessness simply takes over" (p. 284).

This has potentially serious implications in the light of MEXT requirements that there should be a "focus on abilities of teaching staff" (MEXT, 2001) with comments from students being a core component of evaluation. To ensure consequential validity in summative evaluation, students need to realize that their opinions do matter. As Waters (1985) notes, evaluation encourages students to "contribute their ideas which creates an implicit expectation that something can be done to take their ideas into account" (p. 5). In order for students to feel connection with their comments and feedback to faculty, there must be "tangible immediacy to the results" (Spencer and Schmelkin, 2002, p. 397), yet as many students are not taught in subsequent classes by the same teacher, they cannot see any "improvement" and thus feel that expressing opinions has empty benefits. Validity is compromised through such learner indifference. Many students have a cynical attitude which teachers themselves may "inadvertently promote through haphazard or scornful administration" (Smith and Carney, 1990, p. 1), and as the use of results is often unclear to students, evaluation becomes a "perfunctory exercise of little impact" (p. 6), which jeopardizes reliability and consequential validity. Teacher concern over fairness and usefulness of feedback has led to claims that only 23% of faculty made changes to their teaching based on student evaluations (Senior, 1999). This is compounded through questionnaire fatigue; the insistence on the use of standardized ratings which requires that students fill out the same form on every instructor (Abbot et al., 1990). This can lead to a "coat-tails effect" (Block, 1998, p. 418) where a respondent assigns the same number for all questionnaire items out of inertia or a lack of interest, causing
potentially disastrous consequences if results are norm referenced. Many teachers would be surprised that students are empowered to yield so much influence although they often are less than whole-hearted in their contributions. Teachers would be shocked to learn that student evaluation, being simple to collect and analyze, is often the sole criteria for dismissal of part time teachers by the administrators in Japanese universities.

Little is known about the actual processes students follow when responding to rating forms. Do students respond to items by comparing the instructor’s performance to that of other instructors or to some ideal? What are learner motivations to respond? A prerequisite condition for teachers to make improvements to their teaching after student feedback is that they consider student opinion intrinsically worth listening to. However, this appears to be often not the case. As Theall and Franklin (2001, p. 51) have noted, the absence of clear policy, poor instruments and arbitrary decision making has lead to “a great deal of suspicion, anxiety and even hostility towards ratings,” while Scriven (1981, p. 244) asserts that teacher evaluation is a disaster: “The practices are shoddy and the principles unclear.” Attempts at making the student a participant in the evaluative process will continue to founder if both teacher and learner continue to feel that the evaluative process is not fulfilling its ultimate purpose—— the improvement of teaching.

Teacher and learner views on effective teaching

Many writers, for example Feldman (1988), note that if faculty and students do not agree as to what constitutes effective teaching, then faculty members may well be “leery” (p. 291) of students’ overall ratings of them. Teachers believe that students may use different priorities than they themselves would in arriving at overall conclusions. As ratings are on numerical scales, once numbers are assigned faculty committees begin to make comparisons between teachers based on
averages and assume that if one number is larger than another, there is a real
difference between the teachers to whom the numbers have been assigned
(McKeachie, 1997b). Faculty members are supplied with norms indicating the
average ratings on each item. Thus a faculty member whose students all “agree”
that he or she is an excellent teacher, will find that he or she is “below average”
because other faculty members have students who “strongly agree” on that item.

To complicate matters further, reports of other people’s approaches are
themselves value-laden. (Brown, 1993). Effectiveness may be as much concerned
with efficiency as it is with quality of learning. Institutional pressures affect
academic subjects and departments. Some departments value the use of small
group methods and project supervision whereas others prefer large lectures and
written exams, regardless of any external pressures. The espoused values and
values in use may not be consonant within departments.

Research questions

How do learners themselves “rate” the end of semester rating process? If
students question the purpose and see it as a meaningless ritual this will have an
impact on ratings scores of EFL teachers. Problematizing practice needs, as
Pennycook (2001, p. 8) notes, “some sort of vision about what is preferable” to
escape from unwanted labeling of merely “offering a bleak and pessimistic vision
of social relations.” There is also a wider, political question of who defines and
articulates knowledge about teaching and whether teachers should be benign
recipients. It is hoped that this study will create opportunities to discuss and debate
with others involved in the educational process and help to “establish critical
communities of enquirers” (Carr and Kemmis, 1986, p. 40). Critical thinking is
not exercised in a vacuum ; it is developed and refined in active inquiry.

In order to have some idea of what teachers and institutions might want to do
differently, some insight on the following issues needs to be gained:
1) Do learners think evaluations are useful? Why or why not?
2) How should evaluation be subsequently used?
3) What should happen to teachers who receive a low evaluation?
4) Do learners think teachers care about evaluation?
5) What are the important characteristics of effective teachers?

Methods

The author in accordance with school policy administered an official student evaluation form in the thirteenth week of the fifteen-week semester in nine classes of first-year students at a (former) National University in Western Japan. There were around forty students in each class and they were majoring in Medicine, Engineering, and Law. The students were assigned to the classes following a placement test and the classes were compulsory for first year students. After completion, students placed their completed surveys and comments in locked "evaluation boxes” placed around campus, which were emptied by office staff. The week after the official administration, a survey of 24 closed items was administered (see Appendix 1) by the author. The evaluation form was administered separately three times: Twice towards the end of the first semester, and one towards the end of the semester within the same three faculties. Therefore, students in nine classes participated. The students were told that the researcher wished to gain some knowledge about classes views on student evaluation forms, that student responses were completely confidential and that participation was not compulsory. Three hundred and seven students filled in the form. The students were asked to circle the most appropriate answer on a 7-point agreement / disagreement continuum like this:

Please read the questions and put a circle around one number:

Strongly agree 7 6 5 4 3 2 1 Strongly disagree
The quantitative data are shown as percentages in Tables 1 to 5. While the results are from one university in Western Japan, it is hoped that the findings are generalizable to similar contexts.

Subsequently, the theme of effective teaching centering on whether “strictness” in EFL teachers was a positive or negative attribute was explored in the first three classes as part of a scheme of work following the syllabus. Students were asked to list up to five attributes that they believe good teachers hold and eighty–one students volunteered their comments.

The procedure

The data from the closed item survey were analyzed using SPSS v.11.0 for descriptive statistics and an acceptable Alpha reliability score of .79 was obtained. For the purpose of analysis, the responses to ‘Strongly agree’ (7), ‘Agree’ (6), and ‘Slightly agree’ (5) are combined to create an overall score of agreement with the question. Similarly, the sum of responses to ‘Strongly disagree’ (1), ‘Disagree’ (2) and ‘Slightly disagree’ (3) were calculated to gain a measure of disagreement. These aggregated data are reported on in the subsequent analysis of findings. The open–ended data were analyzed using key–word analysis and data were classified according to themes which emerged from the data.

Findings of the closed–item questionnaire
1. Students’ general impressions of the evaluation process

Looking at the data, 75% of students feel that evaluation is a good thing to do, while 71% believe that the teachers somehow do use ratings and comments to improve the class. However, only 21% agree that students are clearly told how evaluations will be used, while 40% feel that the evaluation process could be better. Smith and Carney (1990) similarly found that while students were uncertain of how evaluation was used, students stated that they evaluated their
Table 1: Students’ general impressions. N = 307. Student responses as a percentage.

<table>
<thead>
<tr>
<th>Q1. Evaluation of teachers is a good thing to do</th>
<th>Strongly agree</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>Strongly disagree (1 missing answer)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
<td>Mean: 5.2</td>
<td>S.D: 1.234</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>6</td>
<td>16</td>
<td>33</td>
<td>29</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q2. Teachers use students’ evaluations to improve their classes

<table>
<thead>
<tr>
<th>Q3. Students are clearly told how evaluations will be used</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>15</td>
</tr>
</tbody>
</table>

Q4. The evaluation process could be better

<table>
<thead>
<tr>
<th>Q4. The evaluation process could be better</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

Perhaps this could go someway in dispelling many teachers’ fears that “students too often use the power of their pencils to get even with professors and rating systems may turn the evaluation of effective teaching into a personality contest” (Kulik, 2001, p. 10). Students believe that one purpose is course improvement, but are uncertain whether ratings are used to determine salary, promotion, or teacher retention. Such information needs to be shared so they are aware of the impact of their ratings. This lack of clarity over the usage may come from the haphazard way evaluation is administered reflected in the lack of clear MEXT guidelines outlined earlier. If evaluations are distributed at the very end of the class period with only five minutes for completion only the most frustrated or the most altruistic student will provide more than cursory input, with the teachers themselves communicating that evaluations are an afterthought or something unimportant. Teachers administer evaluations themselves going through the process while not being sure of, and not being in a position to explain the purpose. Teachers often explain the purpose based on their own perceptions or fears in the absence of guidelines. Such instructions can bias student responses.
Some teachers may leave the room while administering which may influence ratings (Seldin, 1993) while Wachtel (1998) states that teacher presence may have a latent effect as the person passing out the forms. For many, evaluation becomes a "ritual that principals and teachers engage in because it is expected—not because they value it" (Good and Mulryan, 1990, p201). Students may question the use the same form in every subject leading to a feeling of decreasing impact each time the evaluation "ritual" comes round. Thus, the use of evaluation by some unknown administrator for some unknown purposes biases the results as the literature suggests that students tend to be lenient if they think data will be used by someone other than the instructor (Cashin, 1988; Centra, 1995) for non-diagnostic purposes. The results of evaluation need to be made available to students as expanding public access could invigorate the process. Teachers would be encouraged to reconsider their own level of investment, while publicity about the purpose and importance of evaluation through campus newsletters, e-mail systems and internet sites could all be used in campaigns highlighting the importance of feedback and reducing the routineness of evaluation.

2. What should happen to teachers with consistently poor evaluations?

In Table 2, 34% of students feel that teachers who consistently get poor evaluations should lose their job, while 41% feel that such teachers should get a salary cut. Eighty-two percent suggest that "effective teacher" training should be introduced which has implications for definitions of what an "effective" teacher does and to whose set of criteria are teachers accountable. Fifty-four percent of participants suggest that the teacher should no longer teach the class if the teacher cannot change the teaching approach to suit the students. Therefore, to increase validity, student awareness of the importance of their input needs to be raised. Studies suggest that students wanted the results made public as not only teachers but students also have the right to view them for future class selection (Spencer
Table 2: What should happen to teachers with consistently poor evaluations. N = 307.

Q5a. The teacher should lose his or her job

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Mean</th>
<th>S.D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>21</td>
<td>24</td>
<td>11</td>
<td>14</td>
<td>9</td>
<td>3.93</td>
<td>1.749</td>
</tr>
</tbody>
</table>

Q5b. The teacher should get a cut in salary

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14</td>
<td>10</td>
<td>16</td>
<td>20</td>
<td>14</td>
<td>15</td>
<td>12</td>
</tr>
</tbody>
</table>

Q5c. The teacher should get “effective teacher” training

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>10</td>
<td>21</td>
<td>24</td>
<td>37</td>
</tr>
</tbody>
</table>

Q5d. The teacher should not be allowed to do the class

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>3</td>
<td>16</td>
<td>20</td>
<td>18</td>
<td>20</td>
<td>16</td>
</tr>
</tbody>
</table>

and Schmelkin, 2002).

Brown and McIntyre (1993) note that different groups have different criteria against which they make their judgements about teachers which include personal achievements or characteristics such as whether the teacher seems to be a warm, encouraging courteous person with a sense of humor. Brown and McIntyre note that there is a :

plethora of criteria used to judge teaching and teachers. All are salient to some group’s concerns for what, in their eyes constitutes ‘good teaching’; but they are of virtually no help in our efforts to understand teaching (p. 23).

Instructional support is crucial: in the face of negative ratings, without support faculty who have prepared long and hard have to reconcile certain knowledge of effort expended against lack of success. If the situation persists over time, these faculty may develop negative attitudes towards students and student ratings, leading to “professorial melancholia” (Theall and Franklin, 2001, p. 47) including hostility towards students and administrators, arrogance, alienation and grade
abuse against students. So that evaluation is not punitive, resources for improvement and support of teaching and teachers must be part of a complete system and cannot be omitted. Support for teaching from administrators and senior teachers is required if the evaluation process is to be perceived as useful and non-threatening. Anything less results in polarized views about the purposes of evaluation which leads to anxiety, resistance and hostility.

3. Confidence in the evaluation process

**Table 3 : Student opinions. N = 307.**

Q6. Teachers don't care about students' evaluations of teaching

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Mean</th>
<th>S.D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7</td>
<td>12</td>
<td>21</td>
<td>26</td>
<td>22</td>
<td>10</td>
<td>3</td>
<td>3.84</td>
<td>1.461</td>
</tr>
</tbody>
</table>

Q7. I think that teachers get friendlier as evaluation day comes near

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>39</td>
<td>23</td>
<td>19</td>
<td>12</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Q8. When I'm evaluating, I'm worried the teacher will somehow recognize my handwriting

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>55</td>
<td>15</td>
<td>11</td>
<td>10</td>
<td>7</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.13</td>
<td>1.557</td>
<td></td>
</tr>
</tbody>
</table>

Thirty-five percent of students feel that teachers do not care about the results of students' evaluations. Universities need to ensure the **tangible immediacy** of evaluation as in many cases students will not have a class with the same teacher again, and so expressing opinions are not beneficial in the long term for students. Dunegan and Hrvinak (2003, p. 284) add that expectancy and equity theories of motivation may provoke mindlessness predicting that students will invest greater effort only when those efforts will produce additional benefits. If students receive virtually no feedback and if there is no strong correlation between positive (or negative) evaluation and teacher rewards (or punishments) then students see very little connection between their evaluations and subsequent outcomes. It must be remembered that the participants in this research were first year students.
encountering evaluation for the first or second time. If evaluation is allowed to become a meaningful task instead of something unique and novel, we are creating a situation where “mindlessness simply takes over” (p. 284).

As Ballantyne (1998) notes, it is important in the teacher / student relationship that teachers are seen as a “real person” (p. 157) who makes errors, but who takes comments on board. Teachers should tell students of changes made due to constructive feedback, or carry feedback over from one semester to the next, announcing at the beginning of a new course that they are trying a new approach based on comments of previous students. Acknowledgment reinforces the importance of ratings and comments. Only 7% of students feel teachers get friendlier in the lessons approaching evaluation. This may mean that students feel teachers are unconcerned or indifferent to the evaluation process and that, again, the ritualistic meaningless belief is emphasized. Similarly, 11% feel that teachers will be able to recognize handwriting, which may bias ratings that are collected often on the same day as summative, end of semester tests.

4. What are the evaluations used for?

In Table 4, 71% of students feel that teachers use evaluations to improve teaching (q.9c) while learner ratings show that only 14% feel that evaluations are used in determining salary, and 24% to make decisions about promotion. Similarly, only 7% think that evaluations are used in teacher retention (q.9e). As noted earlier, if students know that evaluations are used in personnel decisions, they are likely to be more generous in their appraisals, which has obvious implications for consequential validity of ratings.

Students are not aware of the hidden agenda in that they lack information about a crucial purpose of evaluation, the retention or non-retention of teachers. One impression gained from the learners’ answers is of a lack of knowledge about the administrative procedures; students seemingly believe forms are just collected
Table 4: What students think evaluations are used for. N = 307.

<table>
<thead>
<tr>
<th>Q9a. I think they are used to make decisions about teachers’ salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>37</td>
</tr>
</tbody>
</table>

Q9b. I think they are used to make decisions about promotion

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>14</td>
<td>15</td>
<td>18</td>
<td>16</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

Q9c. I think they are used to make course improvements

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
<td>6</td>
<td>17</td>
<td>28</td>
<td>23</td>
<td>20</td>
</tr>
</tbody>
</table>

Q9d. I think they are used to improve teaching

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4</td>
<td>6</td>
<td>20</td>
<td>26</td>
<td>25</td>
<td>18</td>
</tr>
</tbody>
</table>

Q9e. I think they are used to fire teachers

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>16</td>
<td>16</td>
<td>17</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

from boxes around campus and simply handed to the teacher. Teachers are then presumably free to use or disregard as they see fit. It is important to remember the cyclical nature of evaluation, and in order to “close the loop” (Harvey, 2003, p. 4) there is a need to provide feedback of outcomes to students via postings on noticeboards, e-mail or Intranet. As well as providing data on the purpose, how data will be used and disseminated, publication of feedback provides consultation to initiate action in response to students’ concerns, and it shows to students that their comments are valued.

5. What do learners think should be evaluated?

Simmons’ (1996) claim that students may judge a teacher as not “aesthetically acceptable” and are “rendered less capable” of teaching, based on teacher’s physical appearance seem to have been discounted in the present study by 88% of participants. However, 41% of students feel that evaluating the teacher according to how amusing they are is valid while 64% feel they should evaluate teachers by
Table 5: How students think teachers should be evaluated. N = 307.

<p>| Q10a. I think teachers should be evaluated by their personal appearance |
|-------------------|-----|-----|-----|-----|-----|-----|-----|</p>
<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Mean</th>
<th>S.D</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>19</td>
<td>14</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1.91</td>
<td>1.307</td>
</tr>
</tbody>
</table>

Q10b. I think teachers should be evaluated by how much they use AV materials

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>22</td>
<td>18</td>
<td>13</td>
<td>6</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Q10c. I think teachers should be evaluated by how amusing they are

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>11</td>
<td>14</td>
<td>22</td>
<td>22</td>
<td>11</td>
<td>10</td>
</tr>
</tbody>
</table>

Q10d. I think teachers should be evaluated by how much homework they give

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>18</td>
<td>19</td>
<td>22</td>
<td>10</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Q10e. I think teachers should be evaluated by how interesting the textbook is

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>7</td>
<td>9</td>
<td>20</td>
<td>27</td>
<td>18</td>
<td>12</td>
</tr>
</tbody>
</table>

Q10f. I think teachers should be evaluated by how punctual they start and finish class

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>10</td>
<td>16</td>
<td>23</td>
<td>21</td>
<td>12</td>
<td>9</td>
</tr>
</tbody>
</table>

10g. I think teachers should be evaluated by how friendly they are

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>4</td>
<td>7</td>
<td>19</td>
<td>25</td>
<td>25</td>
<td>14</td>
</tr>
</tbody>
</table>

10h. I think teachers should be evaluated by how much they know about the subject

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>10</td>
<td>21</td>
<td>30</td>
<td>34</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean</th>
<th>S.D</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.11</td>
<td>1.674</td>
</tr>
</tbody>
</table>

5.69 | 1.366 |

their friendliness (q.10g). Fifty seven percent of students rated interest in the textbook as important which has huge implications for part-time teachers who often do not have a choice of text and thus teaching material not suited to their teaching approach. Using similar high inference items in end of semester evaluation will lead to an interpretation of competent teachers as only needing good communication skills. Teaching involves more than effective communication as it entails the application of principles of human cognitive
development, understanding of human motivation, as well as pedagogical skills necessary to help students understand the curriculum.

However, 83% of students thought that teachers should be evaluated on subject knowledge, or as Richards (2001) calls it “content knowledge” (p. 209). However, this is just one knowledge base of effective teaching alongside practical knowledge, a repertoire if classroom techniques, and pedagogical knowledge including the “ability to restructure content knowledge for teaching purposes and to plan, adapt and improvise” (ibid.). Arguably, there should be discussion beyond simplistic manifestations of knowledge, as to how different kinds of knowledge, the core components, can benefit the learning experience. It needs to be remembered that student evaluation is more a measure of students’ transient satisfaction than of the teaching and learning. Student evaluation, while important, should never be overemphasized at the expense of many other aspects of the communicative English language program which must also be evaluated.

**Findings of the open ended question**

Eighty-four students were asked to list up to 5 attributes of good teachers. The results are shown in Table 6. Results clearly indicate that students see good teachers in terms of a narrow focus on education processes. These emphasize the act of teaching rather than consequences of teaching especially activities which occur during teaching such as actions which make the English lesson an enjoyable experience through smiling a lot or telling amusing stories. This may rebound on English teachers: Aspinall (2003) interestingly links the use of English words or phrases in Japanese advertising and popular culture as it is perceived to be ‘cool’ or fashionable. This exotic and fashionable image of the English language leads many Japanese students to expect entertainment and fun rather than serious study from English classes and the teacher that attempts the latter may be rated harshly. Delivery is seen as important such as establishing rapport and showing enthusiasm
for the subject. There are very few comments on the products of teaching related to positive changes produced in students in cognitive, affective and psychomotor academic domains. It is not clear what activities differentiate good from poor in promoting student critical thinking, engagement and persistence, and it is not apparent why enthusiasm is important or how it motivates learners.

Not a single student mentioned the products of teaching, for example how much or how well a class has learned or any other aspect of achievement in cognitive domains. Such domains include whether the teacher nurtured student expectations, beliefs and concepts about themselves, or encouraged student attitudes, values and interests towards subject matter or interpersonal skills. Issues of content validity also need to be considered. Are such ratings representative of items from a larger population? Are such criteria equally applicable to a variety of instructional contexts? Rapport may be more critical in a small class than a large one, so ratings may reinforce specific faculty teaching behaviors and their use may constrict teaching styles rather than encouraging a diversity of classroom strategies.

As noted earlier, this has implications for questionnaire design as it may indicate a halo effect operating in a rating that is supposed to measure teacher effectiveness. As Shevlin et al, (2000) suggest, ratings are significantly effected by student perception on a variable that should be unrelated to assessment of ability. Kulik (2001) adds if ratings go up for one variable, they will go up for others as ratings reflect the way students feel as well as the way they think. As d’Apollania and Abrami (1997) and Shevlin et al, (2000) indicate, learners rate features of teaching on one overall, global factor of teacher personality or "charisma." As feelings are often diffuse and unanalytical, it could be that teachers with low ratings, who concentrate on their personality, will find the whole profile of ratings going up.
Table 6: Good teachers... N = 84
have a lot of experience 32
smile a lot 30
make students laugh 29
talk a lot 24
tell funny stories 24
ask students questions 22
use only English in class 12
play games 11
give interesting classes 10
are kind and consistent 10
give easy tests 8
make students do a lot of work 6
are positive and enthusiastic 8
have classes that are easy to understand 5
speak in a clear voice 8
use video or TV 5
know students' feelings 5
praise the students 5
are occasionally strict 5
give good grades 4
use handouts 4
have natural pronunciation 4
communicates well 3
tell information useful in students' lives 3
can speak Japanese 3
give easy classes 3
talk to students in English 3
enjoy teaching 3

Some suggestions for future practice
As the rise of consumerism encouraged the system of student as consumer evaluating teachers, teaching with less market value will lose its attractiveness in the market place. However, using student ratings as a single mechanism of evaluation is deeply flawed. Forms are often handed out at the very end of the
class period of the final class a time when students have a high cognitive load preparing for end of semester tests or presentations. The “last minute” nature of evaluation may also communicate to participants in the evaluation process that such forms are an afterthought or something unimportant.

However, the main disadvantage is the restriction in the range of responses so that the questionnaire actually prompts the kind of answers students can give. Students are unable to express what they are truly feeling because the questions represent what the administrator sees as good teaching. Evaluation focuses on the process rather than the outcomes of teaching. While teacher enthusiasm and appropriate use of blackboard may be useful information for surface improvement of teaching, it is not very useful for addressing the effects of teaching. It does not address what was taught and what was learned, the value of learning, or the effects of teaching upon student learning (Pratt, 1997).

Questionnaires are relative or indicative rather than absolute as they are affected by factors such as timing, student motivation, class size and level, and whether the class is compulsory or an elective. A key issue is one of credibility: the degree to which teaching faculty do rely on a piece of evidence to make a judgment, enhance their understanding or make a decision. This is linked with trustworthiness, which refers to how much users should rely on evidence. If evidence is not reliable, it is not trustworthy and should not be considered credible. Student ratings on closed-item questionably measure one product of instruction: namely, student satisfaction and are influenced by the teacher’s charisma rating. Ratings do not measure directly how much or how well a class has learned or aspects of achievement in the cognitive domain and do not measure the affective products of instruction such as student expectations, beliefs and concepts about themselves, student attitudes, values and interests towards the subject matter. As teachers are subject to highly emotive “evaluation” and its implications for good teaching, to encourage credibility there needs to be a system
whereby teaching is not valued against mandated standards and low inference comments from students. Amongst English teachers, administrators and arguably the students as stakeholder, there needs to be a consensus about what is important in ELT, what will be evaluated, who will contribute and what criteria will be used as an important first step in good practice (Theall and Franklin, 2001). As Seldin (1993) notes, the use of ratings does not automatically lead to improved teaching. A number of factors are at work including whether the comments and numerical ratings are new, fresh or have meaning to the teacher; whether the teacher is personally motivated to improve; and importantly, whether the teacher knows how to improve. Open-ended questions can provide a unique chance for students to make their own points independently, so they deserve to be taken seriously. Some evaluation forms provide space for student comments, but these are not included in feedback and probably not taken into account by administrators in their appraisal. If students do not realize any personal benefits from the process, or if they do not feel their input will lead to major change, we cannot expect learners to be motivated to invest time and cognitive resources. Robust analysis of comments needs to be taken to be alert for patterns and common issues those students address.

Students’ conceptions about good teaching often do not match the conceptions of the instructors themselves with students stressing personality aspects and lectures being entertaining. These differences in student and faculty views might well contribute to the tensions found in some college classrooms and further heighten teacher cynicism. Also problematic in certain school cultures are students who have little intrinsic interest in a course in particular or studying in general and appreciate easy teachers and may rate them higher. Therefore, credibility is threatened when teachers are compared with one another using numerical means or medians. Comparisons of ratings of different classes are also dubious because of different goals, teaching methods, content as well as the conceptions of good
teaching held by student and teacher. Therefore, the extent to which learners are motivated to study inside and outside the class as a measure of learning could be influential, but there is a need to compare instructor’s ratings not with “all” courses, but with similar courses (class size, compulsory or elective courses). Facilitating dialogue between instructor and students can explain products and processes including how learning will be achieved as well as outcomes and gives learners opportunities to treat their own teaching as topic of inquiry. Teachers should encourage students to take ownership of their own answers and comments through dialogue as, without this question, many students make unrealistic comments based on unrealistic perceptions. In general, students may complain about the quality of the lecture but rarely consider whether it was an appropriate method to use.

Scriven (1981) makes the interesting but overlooked point that if evaluation is for development, why is there a need for supervision of the questionnaire or indeed for formative evaluation be done at any specific time in the school semester? If one wishes to get useful information for improvement, students should feel free to submit anonymous suggestions and criticisms at any time through the course. Similarly to heighten expectancy motivation perhaps we should consider abandoning anonymity so that students can feel responsible for evaluation and that their personal opinion will have an impact.

Valid student rating must assess accurately instructor impacts on both processes and products of learning: not only the extent ratings reflect what they do but also extent to which students learn course content, are motivated, and develop critical skills. Greater research on the links between good teaching and good learning would encourage greater assessment of the degree to which student ratings reflect what teachers do (process) and impact teachers have on students (product). We have a situation whereby certain classroom processes (use of AV resources, classes with “fun” content) are regarded as worthy, encouraging a prepackaged
agenda where teachers' purpose is devised by students. This is regardless of whether the subject or teaching approach supported such values, which would then threaten those teachers' futures who were unable to create such a package.

Reliance on student ratings reinforces a conservative pedagogy whereby there is a notion that an ideal teacher exists and that teachers can improve by changing their behavior to more closely match the students' ideal which is often socio-historically determined based on previous learning experiences. If the use of ratings is seen to reinforce specific faculty teaching behaviors, their use may constrict teaching styles rather than encouraging a diversity of classroom strategies. Traditional student ratings may no longer be appropriate to use in the assessment of the quality of teaching, deterring faculty from exploring and using a variety of teaching methods in a teaching–learning environment which encourages students to be passive and teachers to be active. It reinforces the metaphor of the teacher as the expert, the knowledgeable master, transmitting knowledge to ignorant students.

Cashin and Downey (1992) argue for single global items in ratings for personnel decisions as global questions account for a great deal of the variance resulting from a weighted composite of multidimensional student ratings and correlate higher with teacher effectiveness. Students would be more willing to rate carefully items on a long diagnostic form if they knew that long forms were used only when the instructor is working to improve the course. To avoid evaluation becoming meaningless, the number of questions needs to be reduced so it is demonstrable that questions show teaching merit or demerit. Those questions that are not immediately relevant for all classes should be excluded, as there is no purpose in doing personnel evaluation. It is imperative that if students are unaware that evaluations are used for summative purposes including teacher retention and tenure issues that students be informed of them so that they know the impact of their ratings. This may help alleviate incidental fears that their opinion counts for
little.

Conclusion

The consequential validity of student ratings is judged by impact on student learning. Possible inadequacy can be determined by answering the question: how do student ratings influence teaching and organizes the learning environment? Concerning criterion-related validity, are students accurate, consistent and unbiased judges of how well (quality) instructors teach according to dimensions specified on ratings forms? As evaluation has multiple purposes, several procedures are require to expand opportunities for teachers to reflect on instruction by analytically examining classroom processes. There is a tendency to equate limited but important knowledge about one aspect of teaching with effective teaching in general. All parties must recognize that numerous teacher behaviors combine to affect student learning as it is often forgotten that skills instruction must be adapted to subject matter, students, and classroom context factors. Linking teacher behavior to student outcomes does not yield rules or simple answers to issues of teaching and learning.

References


Appendix 1

This is a survey about the end of semester evaluations you complete. I would like to know your opinions on student evaluation. Please do not write your name. Please read the questions and tell me what you think about the teacher evaluation process at University. This information will be used as part of my Doctoral studies at Exeter University in Britain. There are no right or wrong answers. Your answers will remain completely confidential. Thank you!

Peter Burden.

Please read the questions and put a circle around one number:

Strongly agree 7 6 5 4 3 2 1 Strongly disagree

A: Generally speaking:

1) The evaluation of teachers is a useful thing to do
   Strongly agree 7 6 5 4 3 2 1 Strongly disagree

2) Teachers use students’ evaluations to improve their classes
   Strongly agree 7 6 5 4 3 2 1 Strongly disagree

3) Students are clearly told how the evaluations will be used
   Strongly agree 7 6 5 4 3 2 1 Strongly disagree

4) The evaluation process could be better
   Strongly agree 7 6 5 4 3 2 1 Strongly disagree

5) If the teacher consistently gets poor evaluations from students, the teacher should
   a) Lose his or her job
      Strongly agree 7 6 5 4 3 2 1 Strongly disagree
   b) Get a cut in salary
      Strongly agree 7 6 5 4 3 2 1 Strongly disagree
   c) Get “effective teacher” training
      Strongly agree 7 6 5 4 3 2 1 Strongly disagree
   d) Not be allowed to do the class
      Strongly agree 7 6 5 4 3 2 1 Strongly disagree

6) Teachers don’t care about students’ evaluations of teaching
   Strongly agree 7 6 5 4 3 2 1 Strongly disagree
B: About you:

7) I think that teachers get friendlier as evaluation day comes near
   Strongly agree 7  6  5  4  3  2  1  Strongly disagree

8) When I’m evaluating, I’m worried the teacher will somehow recognize my handwriting
   Strongly agree 7  6  5  4  3  2  1  Strongly disagree

9) I think that the student evaluations are used to
   a) Make decisions about teachers’ salary
      Strongly agree 7  6  5  4  3  2  1  Strongly disagree
   b) Make decisions about promotion
      Strongly agree 7  6  5  4  3  2  1  Strongly disagree
   c) Make course improvements
      Strongly agree 7  6  5  4  3  2  1  Strongly disagree
   d) Improve teaching style
      Strongly agree 7  6  5  4  3  2  1  Strongly disagree
   e) Fire teachers
      Strongly agree 7  6  5  4  3  2  1  Strongly disagree

10) I think teachers should be evaluated by
    a) Their personal appearance
       Strongly agree 7  6  5  4  3  2  1  Strongly disagree
    b) How much they use AV materials
       Strongly agree 7  6  5  4  3  2  1  Strongly disagree
    c) How funny they are
       Strongly agree 7  6  5  4  3  2  1  Strongly disagree
    d) How much homework they give
       Strongly agree 7  6  5  4  3  2  1  Strongly disagree
    e) How interesting the textbook is
       Strongly agree 7  6  5  4  3  2  1  Strongly disagree
    f) How punctual they are starting and finishing class
       Strongly agree 7  6  5  4  3  2  1  Strongly disagree
    g) How friendly they are
       Strongly agree 7  6  5  4  3  2  1  Strongly disagree
    h) How much they know about the subject
       Strongly agree 7  6  5  4  3  2  1  Strongly disagree