

TEACHERS' REFLECTIONS ON THE USEFULNESS OF STUDENT EVALUATION OF TEACHING SURVEYS (SETs) AND A MODEL FOR IMPROVEMENT

Peter Burden

(Burden-p@po.osu.ac.jp)

Okayama Shoka University, Japan

Abstract

Sweeping changes in the ways universities are organized and administered worldwide has led to a rise in accountability and the appearance of consumerism as a form of market orientation. In Japan, a results-oriented milieu has been encouraged in response to government demands that the public investment in education justifies closer scrutiny of the outcomes achieved, and for teaching that represents the diversity of students enrolling into universities. With the dramatic decline in the 18 year old population in Japan, a buyer's market has emerged through which student evaluation of teaching surveys (SETs) are used partly as a measure student satisfaction. This is not without debate, and this study seeks to understand the perceptions of 22 local and expatriate English language teachers who participated in in-depth interviews on the usefulness of SETs for teacher improvement. Teachers suggest that using SETs as the sole criterion for evaluating teachers is flawed, unsystematic, and does not lead to improvement, views which often resonate wherever such evaluation is utilized. Participants suggest the need for multi-perspective evaluation which reflects the complexity of teaching and can provide valid data about competence while helping teachers improve the caliber of their work. Thus, a model for improvement based on the 'Stop, Start, Continue' method is proposed which adds context and detail to quantitative data driven issues. Through horizontal dialogue, the implementation of teacher evaluation becomes more efficient and appealing while heightening teacher receptivity.

1 Introduction

The Ministry of Education, Culture, Sports, Science and Technology in Japan (henceforth MEXT, 2004) has advanced SETs as a principal method of getting information from students. This is a reflection of a more economic-centered, more market-sensitive, decentralization movement which emerged at the start of a new millennium in partial response to the expected decline in the 18 year-old population (Yamada, 2001), and the global emphasis on 'quality' in education. As universities' total capacity has equaled the number of applicants, universities are "now subject to a buyer's market where students are courted customers rather than supplicants for admission" (Kitamura, 1997, p. 145). As a higher percentage of high school graduates are entering tertiary education, the content of university education must change to meet diverse students' abilities and knowledge with more consideration required to make education more attractive (Yamamoto, 2005).

1.1 What are SETs?

SETs are student evaluation of teaching surveys or evaluation forms given to students to fill out in classes towards the end of the school semester. These forms often utilize Likert type 1-5 scales anchored from ‘Very poor (1)’ to ‘Very good (5)’ with common questions or factors including specific teacher characteristics about, for example, teacher enthusiasm, way of speaking, receptivity to the students and the use of blackboard and AV devices; whether materials were appropriate and related to, and followed the syllabus; students’ ability to cope with course content, the degree of preparation and reviewing required and the degree of student attendance. These questions are coupled usually, but not always, with a final global characteristic of ‘overall satisfaction’ of the course and ‘effectiveness’ of the instructor. Many schools require the students to anonymously fill in closed item questions twice, with a pencil, once on a machine readable card, which is used for the data analysis by the administration and is the basis for summative scores, and once in paper form. The latter includes an open-ended section for comments which many schools ask the students to fill in. For teachers to read hand-written comments, the administration has to return all the paper evaluation forms to the instructors which many universities are reluctant to do according to the participants. After analyzing the data, school administrators produce a set of quantitative figures for each subject specialty to show MEXT that they are offering quality education to the students.

1.2 What are the functions of SETs?

Hoon, Oliver, Szpakowska, and Newton (2014) suggest that there are three primary functions for the collecting of SETs. Firstly, educators use student-generated feedback to help them improve their teaching as well as to understand the current educational attainment of their students. Secondly, university administrators use student feedback for quality assurance and human resource purposes, while thirdly, prospective students use feedback from previous students to help them make decisions about their education. However, there is still no official policy on timing, administration or explicit statement of purpose delivered either to schools or to teachers. For quality assurance, there is no indication of a remedial path for teachers who receive poor evaluations, often only an internal “symbolically significant attitudinal shift” by university authorities to tie “performance” to salary as experienced by Poole (2005, p. 266) and extrinsic rewards such as the introduction of an awards system, bonuses for “outstanding” teachers, and, conversely, punitive or “appropriate” measures on “incompetent” teachers including teaching suspensions (MEXT, 2001).

In Singapore, the situation appears to be similar in that, ostensibly, the purpose for gaining student feedback is to understand what is working and what can be improved in the teaching programme (Goh & Koh, 2013), with particular attention being paid to student satisfaction. This satisfaction is driven partly by the difficult transition from school to university alongside customer-centric expectations, and the need to reduce attrition or dropping-out which is costly for the school. While views and values of stakeholders may vary, the literature assumes the purposes of evaluation focus either on the formative purpose as diagnostic feedback or the summative

2.1 Data analysis

Subsequently, following transcription of the interviews, the data were “unitized” (Lincoln & Guba, 1985), whereby the text was examined in terms of units of information that became the basis for defining categories. The subcategories were quantified by identifying themes and isolating data by the number of occurrences (Miles & Huberman, 1994).

3 Findings and discussion

3.1 The benefits of SETs

All of the participants accepted that evaluation is necessary as a process to give insights to teachers: all the teachers wished for an open, improvement-focused, cooperative evaluation but, crucially, one that utilizes more than one data source. The results in Table 1 show the benefit of SET teachers gain from their current use in their respective institutions. Teachers talk of feeling “relieved to get positive comments” and that it justifies or reinforces their classroom approach which does not, in itself, encourage reflecting for change. Teachers were prepared to show a commitment for change but did not know how to implement change, as one revealed: “The statistical feedback has just acted as a vote of confidence in my experience. If it were negative, this may be rather difficult to respond to, as there is no detailed link between the problem and the class.” Four teachers noted that getting scores above the average made them feel secure, or glad to get higher scores than their Japanese colleagues.

Table 1. The benefits of SETs feedback

Teacher comments on usefulness	Number of participants mentioning item (N = 22)
Positive evaluations give reinforcement and confidence	10
Teachers can learn if their approach is suitable	8
Teachers can learn if they get better results than others	4
Evaluation process makes teachers more conscientious	3
Total	25

Some find comments about the degree of teacher approachability or availability useful, whether the class followed the syllabus and whether the teacher adhered to class finishing time. Those teachers who did get negative feedback but were prepared to show a commitment for change did not know how to implement change, as one revealed:

“The statistical feedback has just acted as a vote of confidence in my experience. If it were negative, this may be rather difficult to respond to, as there is no detailed link between the problem and the class.”

3.2 The lack of applicability of feedback

Table 2 displays why teachers believe the feedback is difficult to apply in their English teaching situations.

Table 2. Lack of applicability of feedback

Reasons	Number of participants mentioning item (N = 22)
Suspicion over the validity of student responses	20
The late return of feedback	17
Results are difficult to understand or interpret	9
Student comments are too general	8
There is a lack of feedback follow-up	6
Total	60

Many teachers are suspicious whether students complete evaluations in a meaningful way or if they are “simply discharging a boring chore?” (McKeachie, 1997, p.1223). Evaluation is both cognitively taxing and emotionally demanding at the end of term, when students are inundated with concerns over class tests and grades. This adds to the view that evaluation is a meaningless distraction (Dunegan & Hrivnak, 2003). As there is so little information of any worth, and little or no mechanism for remedial help, the potential for teacher growth is limited. Seventeen participants commented on evaluation results being so late they were meaningless while six teachers who had forgotten even having administering SETs were surprised to receive results, unannounced, months later.

Results are too late even to inform the second semester; however, even getting data does not always guarantee it is understandable. Interpreting the data can be challenging and difficult to digest for a busy teacher who cannot read formal Japanese so well. The graphs “look the same” so teachers do not pay attention to them which reinforces the technical, depersonalized top-down nature of evaluation. For eight teachers the comments are seen to be far too general. For students, an ethos of suspicion about purpose can do little to encourage them to give full and frank comments. One participant notes that the “biggest response is the lack of response” receiving an average of five completed surveys out of a class of 25 or 30 as feedback, while another adds that “most of the students didn’t write comments because they didn’t have any interest and just wanted to leave as fast as possible.” A third teacher notes:

“For all the effort the university puts into it I don’t think it’s all that worthwhile. But when I do get the results I get a couple of hints and tips that I can use. But considering the number of questions out there and everybody running around and so on I don’t think the actual information you get is that valid or valuable.”

The role of reflection was seen as of greater value than delayed feedback as some teachers constantly fine tune their classes so timing is all but irrelevant for development.

Six teachers question the utility for development as there is no remedial path or evaluation follow-up which rewards good results. Ratings reinforce minimum standards with little incentive

to seek improvement. If teachers do not receive a high effort reward they will not apply themselves to the evaluation task. The relationship between performance (what a teacher does), evaluation (judging the value of what the teacher does) and rewards (rewarding faculty members in some way) must be recognized but if there are no rewards “evaluation loses its potential and becomes an unnecessary expense in time and effort” (Braskamp, Brandenburg & Ory, 1984, p.20). Teachers also suggested that even if the data were negative, the lack of any detailed link between the problem and a solution meant that evaluation information is judgmental and lacking helpful advice. Therefore, teachers are unable to improve even if they value the information. The evaluation device has to identify particular difficulties instead of just broad areas of concern. As there is no suggestion of the cause it is difficult to suggest how to make changes, and without the chance to consult or to experiment with change, this could easily make evaluation results worse.

3.3 How evaluation can be improved

All of the participants accept evaluation is necessary as a process to give insights to teachers. While teachers held a wide range of views as shown in Table 3, they all wished for a more open, improvement-focused, cooperative, but specific, evaluation. Many participants suggest the need for more teacher involvement, more dialogue between teachers to discuss the results to aid the reflective process for change, and to remove competitive feelings and the pervasive atmosphere of secrecy that surrounds data results.

Table 3. How evaluation could be improved

Teacher comments about improvement	Number of participants mentioning item (N = 22)
Need more feedback discussion	11
Has to be more teacher involvement in decision-making	9
Teachers need to be aware of the impact of evaluation	7
Ratings scales need to be changed	7
Should be linked to students' performance objectives	7
Need to include other evaluation methods	6
Student anonymity needs to be revised	5
The number of questions should be increased	5
Evaluation timing should be changed to mid-semester	4
Question content needs to be revised	4
Need subject specific evaluation forms	4
Explanations to students need to be much clearer	4
Evaluation needs to be less formal	3
Needs to be consequences of evaluation	3
A standard for students needs to be set	3
Evaluation needs to have a clearer framework	3
Repetitive nature of evaluation changed	2
Evaluating body needs to be evaluated	1
Administration also needs improvement	1
Total	77

Feedback should be fast, detailed, and made public while discussion would raise awareness of both teaching styles and why cross-curricular forms are inappropriate. Openness about the process encourages knowledge of both the purpose and what happens to the surveys after they leave the classroom. It should also be made clear how important each student's opinion is, how the opinions impact on non-tenured teachers, and on elective classes. If the university evaluating body has criteria by which the evaluations are reviewed these should be made known; if there is an overall objective to which teachers are supposed to be working it would be useful to know what that is so that classes might be adjusted. While teachers do not wish to take a lot of student time, more specific questions would push students to think more about answers. The questions need to be constantly updated to meet needs of new students and to increase face validity. If there must be an official evaluation teachers should be involved in the preparation of questions, which teachers think would raise responsibility and self-esteem. Teacher involvement would create discussion of the importance of teaching and different styles, creating less of an imbalance towards research and would lead to a more self-reflexive stance.

If ratings are for diagnostic improvement, a more useful process would be for teachers to make their own evaluations to tailor questions to reflect their own classrooms, and administer them during the semester at mid-term so teachers can make use of feedback and students can see change in the lifetime of the class. Teachers, and often administrators, are unaware of the purpose of the evaluation which is not explained and often are just expected to administer without any consultation or input into the questions. Even if the evaluation is intended for formative development, many teachers do not gain any new knowledge as they question the value of the source of information, students' ability to evaluate, and the ability of non-English language teacher specialists to recognize 'good' teaching.

4 Evaluation for teaching improvement

A theme of this research is that a single mechanism of student evaluation has its drawbacks. Feedback through SET is relative or indicative rather than absolute and should use a range of mechanisms using a mix of qualitative and quantitative feedback. A multi-perspective evaluation should reflect the complexity of teaching and can provide valid data about competence while helping teachers improve the caliber of their work. A key element of receptivity or interest in evaluation is the degree to which teachers are able to utilize results without any feelings of threat. Centra (1993) contends that SETs are an example of an evaluative method initially used only for formative purposes when introduced in America. It was only after informative feedback was seen as non-threatening, that administrators endorsed the use for summative purposes. MacBeath and McGlynn (2002, p. 9) suggests that evaluation needs to "cast a wider net than easy to measure performance" and as the value in evaluation is in improving teacher performance, the evaluation must serve as the catalyst for improvement (Stronge, 2006, p.17), requiring a mechanism for communicating both 'why' and 'how' to change.

Through the Stop, Start and Continue method, the students are asked to reflect on four things the teacher should stop doing, things the teacher should start doing and four things the teacher should continue doing. Through such a formative approach, the university can focus more on what the students have to say in their own words and so the survey contents are tailored to what students

themselves think is important (Shah, 2013), as close-ended, quantitative surveys often reflect a teacher-centred or an Administrator’s preconceived framework.

Table 4. “Stop, Start and Continue” formative evaluation method

STOP	List a few things that I (the teacher) do in class that are not working (I should STOP doing): 1 2 3 4
START	List a few things that would be beneficial for me to START doing: 1 2 3 4
CONTINUE	List a few things that I am doing well that I should CONTINUE doing: 1 2 3 4

Gathering feedback at the end of a class in anticipation of revising a subsequent course assumes that student feedback is generic (Cook-Sather, 2009) without recognizing that feedback from learners is context and group specific. If the purpose of an evaluation is to identify how a program can be improved, it is important to ask explicitly what improvements can be made as qualitative feedback from students can add context and detail to quantitative data driven issues (see Hoon et. al., 2014).

Conclusion

“The paradox of teacher evaluation is that it holds the potential to help nearly every teacher improve, yet in actual practice it helps almost no one” state Stiggins and Duke (1988, p. 1). Although almost thirty years have passed since this quote, the paradox remains as SETs seem to have done little for teacher improvement, and that negative consequences are foremost in the minds of many of the survey respondents in this study. If SETs are for diagnostic improvement, a more useful process would be for teachers to create evaluations to tailor questions to reflect their own classrooms, and to administer them during the semester at mid-term so that teachers can make use of feedback and students can see change in the lifetime of the class. Such an example of useful evaluation would be the implementation of the qualitative evaluation method above

because when students choose write positively or negatively about a university learning experience in an open-ended comment it must be of importance to them.

References

- Braskamp, L., Brandenburg, D., & Ory, J. (1984). *Evaluating teaching effectiveness: A practical guide*. Thousand Oaks: Sage.
- Centra, J. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. San Francisco: Jossey-Bass.
- Cook-Sather, A. (2009). From traditional accountability to shared responsibility: The benefits and challenges of student consultants gathering midcourse feedback in college classrooms. *Assessment and Evaluation in Higher Education*, 34, 231-241.
- Dunegan, K., & Hrivnak, M. (2003). Characteristics of mindless teaching evaluations and the moderating effects of image compatibility. *Journal of Management Education*, 27(3), 280-303.
- Goh, K., & Koh, N. (2013). Usefulness of student feedback: The Singapore experience. In C. Nair & M. Mertova (Eds.), *Enhancing learning and teaching through student feedback in social sciences* (pp.111-139). Cambridge: Chandos.
- Hoon, A., Oliver, E., Szpakowska, K., & Newton, P. (2014). Use of the ‘Stop, Start, continue’ method is associated with the production of constructive qualitative feedback by students in higher education. *Assessment and Evaluation in Higher Education*, in press.
- Kitamura, K. (1997). Policy issues in Japanese higher education. *Higher Education*, 34, 141-150.
- Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Newbury Park: Sage.
- MacBeath, J., & McGlynn, A. (2002). *Self evaluation. What's in it for schools?* London: RoutledgeFarmer.
- McKeachie, W. (1997). Student ratings: The validity of use. *American Psychologist*, 52(11), 1218-1225.
- Miles, M., & Huberman, A. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks: Sage.
- Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2001). *Educational reform plan for the 21st Century. The Rainbow Plan*. Retrieved from www.mext.go.jp/english/topics/21plan/010301.htm

- Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2004). *FY2003 white paper on Education, Culture, sports, science and technology*. Retrieved from www.mext.go.jp/english/news/2004/05/04052401.htm
- Poole, G. (2005). Reform of the university English teaching in Japan: A case study. In J. Eades, R. Goodman & Y. Hada (Eds.), *The "Big Bang" in Japanese higher education* (pp. 242-273). Melbourne: Trans Pacific Press.
- Scriven, M. (1981). Summative teacher evaluation. In J. Millman (Ed.), *Handbook of teacher evaluation*, (pp. 244-271). Beverly Hills: Sage.
- Shah, M. (2013). Listening to students' voices to enhance their experience of university. In C. Nair & M. Mertova (Eds.), *Enhancing learning and teaching through student feedback in social sciences* (pp. 29-44). Cambridge: Chandos.
- Stiggins, R., & Duke, D. (1988). *The case for commitment to teacher growth: Research on teacher evaluation*. Albany: State University of New York Press.
- Stronge, J. (2006) Teacher evaluation and school improvement. In J. Stronge (Ed.), *Evaluating teaching: A guide to current thinking and best practice* (pp. 1- 23). Thousand Oaks: Corwin Press.
- Yamada, R. (2001). University reform in the post-massification era in Japan: Analysis of government education policy for the 21st century. *Higher Education Policy*, 14, 277-291.
- Yamamoto, S. (2005). Government and the national universities: Ministerial bureaucrats and dependent universities. In J. Eades, R. Goodman & Y. Hada (Eds.), *The "Big Bang" in Japanese higher education* (pp. 94-105). Melbourne: Trans Pacific Press.