『岡山商大論叢』（岡山商科大学）
第52巻第１号　2016年６月
Journal of OKAYAMA SHOKA UNIVERSITY
Vol.52　No. 1　June 2016

《論　　説》

A comparison of the usefulness for teaching improvement of qualitative feedback data gained from Student Evaluation of Teaching surveys (SETs) and the "Stop, Start, Continue" (SSC) method

Peter Burden

Abstract

Sweeping changes in the ways universities are organized and administered worldwide has led to a rise in accountability where student evaluation of teaching administered through end of semester surveys is widespread in tertiary education. One of the main purposes is for teachers to improve their teaching through utilizing the results of both quantitative and qualitative data produced. However, for teachers to improve their teaching they need to value the new knowledge gained from learner comments. This study compares two methods of gaining qualitative written comments from students in communicative, English language classes; student evaluation of teaching (SET) surveys and a structured pro forma called the "Stop, Start, Continue" (SSC) method. Results garnered from 117 students who completed SET, and 49 who voluntarily responded to SSC, show that the latter received fuller data from students. As students are more engaged in

producing feedback, more abundant, constructive student comments are gained which provides greater insight into strengths and weaknesses which encourages greater teacher reflection necessary for pedagogical improvement.

Introduction

Student evaluation of teaching has become ubiquitous and it is claimed that measuring teacher performance through surveys is used in "in almost every institution of higher education throughout the world" (Spooren, Brockx, & Mortelmans, 2013, p.1). The Ministry of Education, Culture, Sports, Science and Technology in Japan (MEXT, 2004) has advanced Student Evaluation of Teaching (SET) surveys as a principal method of getting information from students. This is a reflection of a more economic-centered, more market-sensitive, decentralization movement for reform at the start of the new millennium because of the global emphasis on 'quality' in education (Leckey & Neill, 2001).

In the literature, there are three primary functions, or pathways, for the collecting of SET surveys (see Richardson, 2005) reflecting different needs, while there remains a "lack of clear guidance on policy and best practice for the most useful feedback from students to assist these different agendas" (Hoon, Oliver, Szpakowska, & Newton, 2015, p.755). Firstly, university administrators need student feedback for quality assurance, accountability, and human resource purposes reflecting the need for "determining competence of teachers in order to assure that services delivered are safe and effective" (Stronge, 2006, p.4) and are seen as being summative. Secondly, teachers use student-generated feedback diagnostically to help them improve, or innovate, their teaching. Thirdly, prospective students use feedback from previous students to help them make decisions about the selection of courses and teachers. A fourth purpose, frequently noted in Asia and Australia, (Goh & Koh, 2013), is the need to reduce student attrition which is costly for the school in the difficult transition from

school to university so evaluation serves to fulfill customer-centric expectations. As the numbers of prospective students decline, more diverse, less traditionally competent, and less homogeneously skilled young people are entering university, who may display "regressive attitudes towards learning" (Tsurata, 2003, p. 131). 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).

Therefore, evaluation has both an accountability-oriented function contributing to the mission of the program, the school and the total educational organization, but it should also be improvement-oriented, contributing to the personal and professional development needs of the individual teacher. Blair and Noel (2014, p.879) note that assessment by students "can be a driver of improvement," but getting the balance between school goals and individual teacher professional growth and improvement is very difficult (Stronge, 2006).

Due to the ease of collection and publication as statistical data, and because "many individuals at various levels of decision making [are] lost without numbers" (Svinicki, 2001, p.17), student of evaluation of teaching is carried out with surveys or ratings forms using fix-ended questions 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. The scores on these questions are then used to generate a statistical report for summative, administrative purposes. These questions are coupled usually, but not always, with a final global characteristic of 'overall satisfaction' of the course and 'effectiveness' of the instructor. The form includes an open-ended section (often on the back of the form) for comments to add context and detail to issues which arise in the quantitative data to inform

teaching.

Open-ended, qualitative data allows students to provide written comments to explain the scores that they assign for closed-ended items and to draw attention to topics that were not addressed in the closed-ended part of the form (Nasser & Fresco, 2002), or to identify reasons for statistical results which may be different from researcher assumptions (Grebennikov & Shah, 2013), or because they represent researchers' preconceived framework, by allowing students a greater freedom of expression. Arguably, the fixed-ended questions explicitly follow assumptions of a teaching method, so even if the evaluation is intended for formative development, many teachers do not gain any new knowledge as they question the value of the information. As Prabhu (1990) notes, if one method is seen as superior, then it is expected that all classroom learners would benefit from exposure to that method, and that by following the procedures there is a prediction of results. Therefore teaching is reduced to a "faithful following of highly specified routine - something of a pedagogic ritual" (Prabhu, 1990, p.171).

Richardson (2005) shows that while evaluations of the same teachers given by successive cohorts are highly stable over time - suggesting that results can be seen as valid - perhaps teachers' performance do not improve, or change, with experience. Results also suggest that teachers gain little new knowledge from data drawn from summative, fixed-ended questions, while scaled items give only scant diagnostic feedback to identify strengths or weaknesses that teachers can remedy (Kember & Leung, 2009).

Ideally, the qualitative and quantitative data should complement each other (Grebennikov & Shah, 2013). Therefore, qualitative data from students' comments can provide useful insight into aspects of courses that learners find important. Learners can specifically comment on teaching aspects that are only generally touched on or measured in generic SET instruments. In previous research, studies of SETs have focused almost exclusively on quantitative SETs data for summative

accountability purposes, while students' comments for a formative, improvement purpose have been largely overlooked. In English language teaching (ELT) it is questionable whether evaluation questions fairly represent the complexity of communicative teaching activities and thus are seen to have 'utility,' or whether there is an implicit single source methodology in SETs.

If an assumption is that teachers use a didactic, lecture style, then evaluation questions may ''provide only a sketchy partial representation of the language course they attend" (Block, 1998, p.150). For example, it has been claimed that within tertiary education there is only a ''fragmented knowledge of CLT" (Sato, 2002, p.45) suggesting a path or one set of discoveries reinforcing a teacher-dependent view of learning as a collection of discrete skills.

Evaluation for teaching improvement

This study considers the perspective of student evaluation for teaching improvement. Such evaluation should reflect the complexity of teaching and provide valid data about competence while helping teachers improve the caliber of their work. As noted, open-ended qualitative data allows students to provide written comments to explain the scores that they assign for closed-ended items, and to draw attention to topics that were not addressed in the closed-ended part of the form (Nasser & Fresco, 2002). Two key words might be "utility" (Stronge (2006, p.9) whereby useful, informative, timely and influential information is provided to ensure that findings are valid and reliable, and that evaluation has "impact" (Stiggins & Duke, 1988, p.58) so that evaluation is seen as an opportunity where the impact is dictated by both the nature and quality of the evaluation procedures and the quality of the feedback provided. However, the quality of data is determined by the appropriateness of the procedures used to gather it.

Open-ended data must serve as the catalyst for improvement, and require a

mechanism for communicating both ‘why’ and ‘how’ to change. Centra (1993) contends that truly significant improvement is likely to take place only if the evaluation fulfills four conditions:

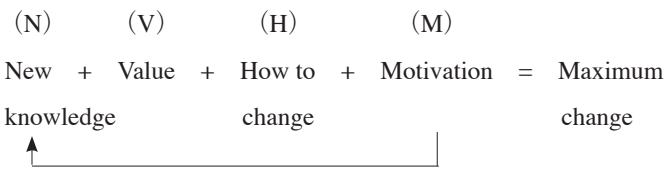

Fig. 1: The NVHM Model for change (Centra, 1993)

In the linear model, when evaluations are used maximum change occurs when teachers receive new knowledge, value that knowledge, know how or receive practical help to change and have the motivation to change. The loop signifies that teachers are motivated to seek new knowledge which they value as a further spur to change. One key principle for teachers is how much teacher growth is engendered in evaluation. For improvement where the formative focus is on diagnostic feedback, teachers need credible data as potential for growth. Brockx, Van Roy, and Mortelmans (2012) suggest that students should be a “commentator” (p.69) on aspects of teaching and courses that students find important.

As an alternative to SET surveys, the ‘Stop, Start and Continue’ (SSC) method has been widely used in the business world (see Steinbrecher, 1992; Procter & Gamble, n.d) but seems to be under-utilized in feedback in tertiary education. In this form of open-ended data collection, students are asked to reflect on things the teacher should stop doing, things the teacher should start doing and things the teacher should continue doing (see the English and Japanese examples used in this study in the Appendix). In one of the few academic studies on the use of SSC, Hoon et al. (2015) conducted a comparison study of the effectiveness of SSC compared with a pro forma, free text entry box form (a blank paper) and the feedback form asking the students to reflect on the three areas outlined above.

The findings of the study found that open, free text responses had large quantities of descriptive feedback that did not necessarily allow for course development while SSC had less feedback in terms of quantity of statements, but with highly constructive feedback in those statements given.

As research into the use of SSC in English language education has not been carried out, the purpose of the present study is to compare two methods of gaining qualitative data from tertiary students studying English as a compulsory subject designed to improve English language teaching. As noted earlier, the most widespread method of gaining student-driven, open-ended data is from questions at the end of a SET survey administered at the end of the semester, while the second method is the 'Stop, Start and Continue' (SSC) method of gaining qualitative feedback from learners.

Research purpose

The author became interested in SSC as part of ongoing research into student evaluation of teaching and in particular after reading George and Cowan's (1999) handbook of formative evaluation techniques. Overall, there is a lack of published research either on the quality of student feedback or on student comments, especially in a language learning context. While universities collect huge amounts of quantitative data, little research has been conducted into what students offer in terms of qualitative feedback. As Hoon et al. (2015) did not compare the effectiveness of SET and SSC, this research compares data gained from 4 classes who filled out both SSC and SET surveys, in an attempt to discover whether feedback generated from the SSC would yield more useful data that teachers can utilize for teacher improvement than data provided by the open-ended comments on SET surveys.

Method

1) The SET survey administration

The official SET survey was administered in 4 "Intermediate English" English as a Foreign Language (EFL) classes taught by the author in which 117 students attended on the day. The administration was on a pre-determined day and class in week 13 of a 16-week semester at a private university in Western Japan. The students were majoring in Law, Commerce, or Business Management studying English as a compulsory subject. As the institution requires SET to be administered in only one class of each teacher's class load, collection therefore took place over 4 semesters. Presumably, this is reduce student overload of having to fill in the SET survey in all the classes, but the chances of being required to fill in the survey in many classes is still high. To encourage the maximum responses, the author elected the Intermediate English class with the most students for the SET and SSC administration in each semester. The classes were not streamed according to ability levels, and a 'communicative' teaching approach was utilized where learning of English is encouraged by negotiation of meaning through pair or group work which often involved information gap activities with collaborative, scaffolded, interaction. This teaching approach is a contrast to what many students experienced in high school from which the participants have recently graduated (see Burden, 2009). Japanese high school teachers "feel the need to primarily conduct teacher-fronted non-communicative activities" (Sakui 2004, p.158), with teachers describing how they prepare students for "grammar-skewed" (159) entrance exams with targeted grammatical features. The implication of the negative washback of entrance-exam centered English is that many students at university expect teachers to use grammar translation pedagogy experienced in high schools. O'Donnell (2003, p.63), for example, found that learning strategies used in university "continue to parallel many of the traditional practices of their secondary school experiences", and Matsuura, Chiba and Hilderbrandt (2001)

found that most students rely on translation believing that translation into Japanese is necessary.

In the SET survey students were required to agree or disagree with twenty statements utilizing a Likert type scale anchored from 1 to 5. On the back of the computer-readable quantitative data the students were asked to fill in with pencil their responses to three open-ended questions which asked the students to say what was 'good', 'unsatisfactory' and any 'opinions for change' about the course. The students filled in the evaluation form during the class time so each student returned a filled-in form. While all 117 students responded to the Likert scale questions, 120 students left the open-ended questions completely blank, meaning that 43 students responded to the qualitative data.

2) The SSC administration

In week 14, 117 students attending the same 4 classes over 4 semesters were asked to take home and complete the "Stop, Start, Continue" evaluation form. As this survey was administered for research purposes rather than for institutional enhancement purposes, participation was voluntary, and un-coerced in accordance with ethical practice. The form comprised English and Japanese language versions, and students could freely choose either version. Filled in responses were received from 49 students, of which only one responded to the English language version. While the response rate is low, this is not unusual. Kember and Leung (2009) reported a return rate of 53.2%, while Stewart (2015, p.5) distributed an online survey with a return rate of 31.6%, of which 44% "opted to leave a comment" representing only 13.9% of the population. Spooren and Van Loon (2012) in a study of 895 students over 24 courses found a response rate of 26.8%. Similarly, Brockx, Van Roy, and Mortelmans (2012) quoted from a study in which the percentage of students who made the effort to write comments on SETs was around 10-12%. As the purpose of the present is to explore dimensions of

feedback language rather than to generalize findings beyond the sample, it can be argued that the data returned is manageable for the purpose required.

Data analysis

After the data were translated into English, it was "unitized" whereby student responses were analyzed to reveal patterns in the data using a key word analysis with categories generated by the statements made by the students. A content analysis of the responses was carried out and the results totaled and displayed in tables to "see the general drift" of the data by showing distributions (Miles & Huberman, 1994, p.253). Similar to Stewart (2015), and encouraged, perhaps, by the format of the survey (see Appendix) the students often responded with single words, typically short phrases, and comments often lacked the sentence structure of traditional formal writing. Categorization was reasonably straightforward following the "constant comparative method" (Lincoln & Guba, 1985) whereby through coding, all the comments for both question were ordered into topics which were then checked across categories for overlapping themes or different nuances of meaning.

Results and discussion

As the SET survey administration was a compulsory requirement, each student in the class completed the form, but the author decided that it was ethical that participants should be free to withdraw from research at any point (Richardson, 2005). While the number of responses received in the SSC administration appears disappointing, it should also be remembered that many of the 117 students chose not to respond at all to the qualitative data on SETs. As the data collection was longitudinal over 4 semesters, the inevitable delay on the return of the SET survey statistical data due to the summative, statistical purposes of this evaluation process did not affect the study. The data was returned two months later, well

beyond the lifetime of the class and also late to inform the subsequent iteration. The data analysis of the SSC was time consuming-which partly accounts for the Administration preferring the organizational constraints of easily countable, quantifiable data.

Table 1 below shows the student responses to the qualitative, open-ended data on SSC and SETs. The 117 respondents of the SET survey made 80 comments in total, while the 49 respondents out of 117 who accepted the invitation to write their views on the voluntary SSC made 137 comments. Arguably this data alone is encouraging and supports Hoon et al. (2015) findings that the SSC provided fuller data than other open-ended methods.

Table 1: Comparison of the number of responses to qualitative data on the two questionnaires

	"Stop, Start, Continue" n=49*	Student Evaluation of Teaching survey n=117	
The teacher should stop doing	24	23	What the student thinks is Unsatisfactory
The teacher should start doing	35	18	Any opinions or suggestions about the class
The teacher should continue doing	78	39	What the student thinks is good about the class
Total number of responses	137	80	Total number of responses

* 117 "Stop, Start, Continue" surveys were administered, and responses received from 49 students

Table 2 breaks down the data further and shows the 23 comments made on the "unsatisfactory" data of the SET survey and the 24 comments on what the teacher should "stop doing" on the SSC. What is apparent is that the number of students responses of "nothing", meaning "no opinion" is proportionally fewer with the SSC and that the data is arguably more meaningful.

Table 2: Comparison of what the teacher should stop doing, or was unsatisfactory

	"Stop, Start, Continue" n=49*		Student Evaluation of Teaching survey n=117		
Stop	Nothing	12	Nothing	17	Unsatisfactory
	Class too fast/fast, difficult	4	T talks English too much	1	
	Not using Japanese/Speaking English so much	2	Homework in English	1	
	Shadowing	1	No Japanese translation of textbook	1	
	Number of times reading conversations	1	Speed is fast	1	
	Small class system	1	The class is hard for shy	1	
	Difficult to understand explanation	1	Don't understand	1	
	The class is good as now	1			
	Writing on the board is difficult to read	1			
Total		24		23	Total

The results of Table 3 below again show that the SSC yielded fuller data, with 35 comments received compared with 18 on the SET survey. It should also be noted that 12 of the 18 respondents replied with only "nothing" to the elicited opinion or suggestions for improvement in the SET, while the similar theme seeking insight into what the teacher should start doing gained fewer replies of "nothing", 9, and a much larger range of insights on the SSC.

Table 3: Comparison of what the teacher should start doing, or opinions and suggestions

	"Stop, Start, Continue" n=49*		Student Evaluation of Teaching survey n=88	
Start	Nothing	9	Nothing	12	Opinions or suggestions
	Small test	5	Homework in English	1	
	Use more Japanese/translation	2	Homework weighting reduced	1	
	Write sentences in English	2	Students asking questions in English in pair work	1	
	Japanese translation of text	1	Class is good as it is	1	
	Easier English	1	Content is difficult	1	
	Increase the homework	1	Teacher asked students from X department more questions	1	
	Use the screen	1			
	More listening	1			
	Make students prepare vocab.	1			
	Increase the question time	1			
	Say the day and date at start	1			
	Speech in English	1			
	Increase student number	1			
	Tourism English	1			
	Use music	1			
	More pronunciation practice	1			
	Homework rechecking	1			
	Self introductions	1			
	More explanation of vocabulary	1			
	Unit new vocabulary introduction	1			
Total	35		18		Total

Finally, in response to what was 'good' in the class, there were 39 comments from 88 respondents in the SET survey, while in the SSC, the 49 students gave 78 comments related to what the teacher should 'continue' doing.

Table 4: Comparison of what the teacher should continue doing, or what was good in class

	"Stop, Start, Continue" n=49*		Student Evaluation of Teaching survey n=88		
Continue	Warm-up cards	13	Pair work with other students	6	Good
	Asking questions in pairs	10	Content is repeated many times, so easy to understand	6	
	Greetings in English	9	Fun class	6	
	Nothing	7	Nothing	4	
	Finding things in common	5	Easy to understand	4	
	Homework	5	Explanations clear	3	
	Ensuring all students participate	4	Foreign teacher so English is standard	1	
	Little grammar, lot of conversation	4	Natural pronunciation	1	
	Walking around the class and talking to other students	4	Became a little interested	1	
	Reading conversations aloud	3	I could participate	1	
	Pronunciation practice	3	Chances to speak	1	
	Whole class conversations	2	Teacher uses big voice	1	
	Choosing students at random	2	Useful as conv. class	1	
	Nothing I want to change	1	Speaking is central so few sleepers	1	
	Everything	1	We can practice one unit over a few weeks	1	
	Writing sentences in English	1	Teacher chooses who will respond	1	
	Dictionary work for unknown words	1			
	Writing lesson content on the board	1			
	Using name-cards	1			
Total		78		39	Total

The data displayed in Tables 2, 3 and 4 are relevant to the author as feedback on teaching in his particular situation and relate to his classroom teaching practice and communicative approach. Results are therefore not directly related to each reader's situation, but are displayed to show how each teacher can gain useful insight to aid reflective practice whenever formative feedback is gained.

Discussion of findings

The results of the SSC are encouraging because there were twice as many written comments from this evaluation event. As participation was voluntary, it

can be assumed that students were more willing to give constructive feedback, there were fewer blank comments boxes, and fewer single word replies of "nothing", meaning they had no opinion, insight, or motivation to contribute.

The comments relate to the author's classes and so the data is not meaningful beyond the class, but there is sufficient data to encourage critical self-reflection, to encourage a "movement towards a self-judgment couched in terms of 'How well do I do it?'" (George & Cowan, 1999, p.2). However, while caution should be taken not to over-interpret the comments as they are perhaps not representative of the whole class, they have value as reflection on practice or "performance" (Schön, 1983, p.61). A classroom teaching can become "repetitive and routine," the opportunity to reflect may be missed leading to teachers becoming "selectively inattentive to phenomena that do not fit" (p.61) perceptions of their own knowledge.

Instead, teachers need to learn more about student perceptions both to affirm areas of good practice, and suggest areas to build on as part of the cycle of continuous reflection and to put to good use as much insight into and knowledge of our students' learning experience. While the inevitable delay of over two months in the return of SETs data can do little to inform reflective teaching as the administration is long past, the SSC data above informed several significant changes in the module design and delivery prior to the arrival of the next cohort.

How to use the SSC data

The SSC was administered after SETs in week 14 for research purposes but, ideally, it needs to be earlier so that feedback should be shared so that students can see the results of their action implemented, and makes the process "worthwhile" as students need to realize "personal benefits" in investing their input (Dunegan & Hrivnak, 2003, p.282). Cook-Sather (2009, p.8) notes that "it can be a challenge to genuinely open up dialogue with students and respond in ways that

both validate the students' input and honor the faculty member's pedagogical commitments" but it is necessary to show the data, such as in a PowerPoint to say "Here's what you are all saying." This is especially important as some comments regarding translation or an increased use of grammar (see Table 3) for example, do not follow the author's/researcher's assumptions about the tenets of 'communicative' teaching.

Displaying data can show where the students are in the class in relation to each other, to realize that different students have different views about their classes and can point out that there are conflicting recommendations such as "use more/less Japanese in class." Displaying data encourages students to understand that other students have different needs to be addressed. As an example from the data above, one student felt compelled to comment in both the SET and SSC that the teacher was "unbalanced" in the selection of which students were called upon to respond (see Table 2). While the author certainly reflected on this, it was realized that this is only one student's view but which could be opened up to the class for comment.

Also, engaging in dialogue can help to clarification of some of the tenets of the communicative approach as learners should not be regarded as passive receptors of lesson content who do not question the reasons for undertaking a task. Instead, there is a need to let the students "into the picture" (Nunan, 1989, p.184) as learners come to class with their own sets of learning objectives and beliefs, which, being socio-historically constructed, may not match the objectives of either the syllabus or the individual class. Teachers need to 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 otherwise the evaluation process becomes a ritual that administrators and teachers engage in because it is expected--not because it is valued.

However, there are tensions between the teacher and the student view of

teaching as "students may not necessarily know what was best for them" (Yao & Grady, 2005, p.123). Teachers can be open to student feedback about mechanical aspects of their teaching, but show reluctance to change aspects of their teaching that they valued over a long period of time such as the teaching style and the course content. So teachers need to be wary of reactions as it is "all too easy to feel, or worse still, to display anger or disappointment when you hear or read a negative message." (George & Cowan, 1999, p.14) as comments could "rupture your ego" (Yao & Grady, 2005, p.123) so it is important not to five unbalanced weight especially to isolated comments that are maybe unrepresentative of the class. Equally, if the evaluation came out reasonably well it is tempting to settle placidly for the decision that no action is needed.

Limitations of the study

Grebennikov and Shah (2013, p.615) note that "anecdotal evidence suggests that many students write comments when they are either quite happy or quite frustrated about something." While studies indicate that closed-ended survey items may not cover issues that are really important for students, educators should realize that if students choose to write positively or negatively about a learning experience in an open-ended comment it must be of importance to them (Shah, 2013). However, the lack of response in the SSC and, arguably more so on the SET is a cause of concern.

The 49 voluntarily completed SSC forms handed to 117 students generated 137 responses, while the compulsory SETs administered to 117 students garnered 80 responses - with many students declined the invitation to comment on the open-ended questions at all - so while it is clear that the data from SSC is fuller, the study can still be seen to have limitations. The students were invited to leave comments on the SSC but close to 50% chose not to do so. As noted earlier, the author considered voluntary participation to be ethical, but instead of institutional

research, making SSC part of the formal teaching and learning process could increase student responses (see Richardson, 2005). Conversely, though, students may feel that it is more of an Administration demand and less of a response to an individual teacher.

Some Implications for future practice

1. The timing of evaluation

One key area to consider is the timing. As evaluation is at the end of the semester it is seen to lack "tangible immediacy" (Spencer & Schmelkin, 2002, p.406), as expressing opinions does not in the long run benefit individual students which may according to motivational theories of expectancy (see Chen & Hoshower, 2003) lead to perfunctory student response and low valence in the evaluation system. Perfunctory, or even lazy, responses affect the validity of feedback. As Svincki (2001, p.18) adds, the students have seen enough examples of "poor teaching" (hopefully not in the author's class) to "make them skeptical about whether anyone actually reads the data," so students many conclude their efforts are not worth putting forward.

Cashin (1988) reported that students are more likely to rate teachers highly on quantitative data if they had a prior interest in the subject matter or were taking the course as an elective, with higher ratings achieved from students who took a course for general interest, and it may be that the students in this study who often saw themselves as elementary or low intermediate users of English even after six years of studying in high school, might not be motivated to participate in writing comments in yet another compulsory class.

2. The repetitive nature of evaluation

The poor quality data gained from the three SET open-ended questions may be due to participants believing that the repetitive nature of evaluation has diminished

the potential for new insights. As teachers only submit evaluation in one class, it would appear that students are not overburdened to participate as in Spooren and Van Loon's (2012) study, but the students may be unlucky enough to find themselves in the sole class that many teachers choose. As Spencer and Schmelkin (2002, p.406) note, "since students are unsure whether their opinions matter, or to what purpose the ratings are put, they may not pay attention to them" which is in contrast to their wish to provide feedback, while helping reduce students' tendency to evaluate only those activities occurring near the administration of SETs. When students encounter something unique or novel they tend to approach it mindfully, but because it is customary for students to complete generic, one size-fits-all SET surveys in many classes mean the procedure has inadvertently become standardized to the extent that "mindlessness simply takes over" (Dunegan & Hrivnak, 2003, p.284).

3. The lack of understanding of the process of evaluation

Another key area is that learners many lack understanding of either the course or the process of evaluation itself. As Nunan (1989, p.3) suggests, it is "unrealistic to expect learners who have never experienced a particular approach to express an opinion about it." The classes in this study follow tenets of communicative language teaching which the students might be unsure upon entry to university after 6 years of high school English education. Students after only a semester may find it difficult to evaluate comparing the narrow parameters and assumptions about classroom teaching and learning implicit in the questionnaires and the classroom ELT experience. There needs to be participation by teachers in negotiating what counts as good practice, which draws on cultural discussions.

Also, comments tend to come "from either the very satisfied or the very dissatisfied" (Lewis, 2001, p.31), if the perceived levels of English ability are considered again whereby 68 perceived themselves as "Elementary" or "Lower-

intermediate," it may be that students do not know enough about English learning or the new approach to make a judgement. Svinicki (2001, p.18) has noted, motivation theory says that if someone does not think he or she can successfully accomplish a task, motivation to engage falls. In this case, faced with the request for feedback and a lack of clear understanding about how to give it, students may choose to say nothing at all or make very general comments that could not be criticized.

4. The assumptions of teaching method

Underpinning the closed data-driven questions on SETs is an assumption of teaching method, often lecture based or a prescribed set of teaching skills. Open-ended comments complements quantitative data and address aspects of the learning experience important to students that are different from the university perceptions so it is vital to incorporate such feedback into the improvement-driven focus. Therefore, we need to consider how we can encourage more students to participate in qualitative data collection.

5. Knowing how to evaluate

Students need to know *how* to evaluate. There are very few opportunities for students to learn the skill of giving feedback to be "more sophisticated evaluators" (McKeachie, 1997, p.1223). If students are more satisfied with opinions collected during the lifetime of the course, maybe teachers should encourage small group discussions with a "class facilitator" to discuss the strengths of the class, areas for change and importantly, how students recommend changes to be made with the caveat that there is a tension at times between validating the students' input and honoring pedagogical commitments. Students could be encouraged to keep a diary so they are "sensitized to [the] recency effect" (Dickey & Pearson, 2005, p.8) to promote a source of stable course evaluations. A "One-minute paper"

(Svinicki, 2001) encourages students to give quality feedback, while students have little opportunity to learn the skill of giving feedback and never receive feedback on *their feedback*. Students spend just a minute commenting on what helps them learn in that class, give views on the content, and outline difficulties or confusion experienced that week. Auerbach (2001) suggests eliciting dialogue by using photographs or video recordings of different learning contexts. Students can see different ways of learning and expand their view of what counts as useful education.

Conclusion: Does more data lead to better data?

Through constant reflection students will question what is beneficial and what has not helped in their own learning. They thus become more critical observers of their own learning. Auerbach (2001) suggests that issues important to students will emerge more readily in an atmosphere where they feel a sense of ownership, and discussion with learners to sensitize them to evaluating their own learning and the conditions that contribute to learning are important in developing their ability to learn more effectively. Of crucial import is that the teacher can encourage comments through early feedback which improves later feedback as students learn that their comments are taken on board and that feedback does make a difference.

One important conclusion is the realization of the need to shift from summative, end of semester to a more formative administration. If evaluation is left to the end of a course it loses any opportunity to inform and influence teaching and so evaluation should take place around mid-term to initiate actions in response to student concerns within the lifetime of the course. Many would agree with Seldin (1993, p.43) that a "measure of student opinion can hardly be ignored: the opinions of those who eat the dinner should be considered if we want to know how it tastes," all the stakeholders in educational evaluation should also note Seldin's caveat that evaluation must be "systematic" and carried out in a

"congenial environment." As noted earlier, maximum change in teaching occurs when teachers receive new knowledge, value that knowledge, know how or receive practical help to change and have the motivation to change. This study found that SSC produces fuller, more informative data than SETs that has "impact" and utility" for teachers providing instructors with insights into the strengths and weaknesses of their teaching practices. This is important as teachers need to reflect and reconceptualize their teaching and without useful feedback fundamental change in practice is unlikely. To foster development, the use of evaluation should improve the quality of teaching through preparing teachers to teach; should provide an environment where they can teach; and, most importantly, should motivate them to teach.

References

Auerbach, E. (2001). Creating participatory learning communities: Paradoxes and possibilities. In J. Hall & W. Eggington (Eds.), *The sociopolitics of English language teaching* (pp.143-165). Clevedon: Multilingual Matters.

Blair, E. & Noel, K. (2014). Improving higher education practice through student evaluation systems: is the student voice being heard? *Assessment and Evaluation in Higher Education*, 39, 879-894.

Block, D. (1998). Tales of a language learner. *Language Teaching Research*, 2 (2), 148-172.

Brockx, B, Van Roy, K., & Mortelmans, D. (2012). The student as a commentator: Student comments in student evaluation of teaching. *Procedia-Social and Behavioral Sciences*, 69, 1122-1133.

Burden, P. (2009). English language teaching: A disadvantaged profession in Japanese tertiary education? In H. Chen & K. Cruickshank (Eds.), *Making a difference: Challenges for applied linguistics* (pp.262-283). Newcastle: Cambridge Scholars Press.

Cashin, W. (1988). Student ratings of teachers: A summary of the research. *IDEA paper No. 20. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.* 1-6

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

Chen, Y., & Hoshower, L. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment and Evaluation in Higher Education*, 28, 71-89.

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.

Dickey, D., & Pearson, C. (2005). Recency effect in college student course evaluations. *Practical Assessment, Research and Evaluation*, 10. Retrieved from http://pareonline.net/getvn.asp?v=10&n=6 December 3 2007.

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

George, J., & Cowan, J. (1999). *A handbook of techniques for formative evaluation: Mapping the student's learning experience*. London: RoutledgeFalmer.

Goh, K., and Koh, N. (2013). Usefulness of student feedback: the Singapore experience. In Nair, C and Mertova, M. (Eds.), *Enhancing learning and teaching through student feedback in social sciences* (pp.111-139). Cambridge: Chandos.

Grebennikov, L., & Shah, M. (2013). Student voice: Using qualitative feedback from students to enhance their university experience. *Teaching in Higher Education*, 18 (6), 606-618.

Hoon, A., Oliver, E., Szpakowska, K., & Newton, P. (2015). 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*, 40, 755-767.

Kember, D, & Leung, D. (2009). Development of a questionnaire for assessing students' perceptions of the teaching and learning environment and its use in quality assurance. *Learning Environments Research*, 12, 15-29.

Leckey, J., & Neill, N. (2001). Quantifying quality: The importance of student feedback. *Quality in Higher Education*, 7, 19-33.

Lewis, K. (2001). Making sense of student written comments. In K. Lewis (Ed.), *Techniques and strategies for interpreting student evaluations*. (p.25-32). San Francisco: Jossey-Bass.

Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Newbury Park: Sage.

Matsuura, H., R. Chiba & P. Hilderbrandt. (2001). Beliefs about learning and teaching communicative English in Japan. *JALT Journal*, 23, 70-89.

McKeachie, W. (1997). Student ratings: The validity of use. *American Psychologist*, 52, 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). (2004). *FY2003 white paper on Education, Culture, sports, science and technology*. Retrieved from www.mext.go.jp/english/news/2004/05/04052401.htm

Nasser, F., & Fresko, B. (2002). Faculty views of student evaluation of college teaching. *Assessment and Evaluation in Higher Education*. 27, 187-196.

Nunan, D. (1989). Hidden agendas: The role of the learner in programme implementation. In R. Johnson (Ed.), *The second language curriculum* (pp.176-186). Cambridge: Cambridge University Press.

O'Donnell, K. (2003). Uncovering first year students' language learning experiences, their attitudes and motivations in a context of change at the tertiary level of education. *JALT Journal*, 25, 31-63.

Prabhu, N. (1990). There is no best method-why? *TESOL Quarterly,* 24, 161-176.

Procter & Gamble, Inc. (Undated). 360° / Stop-start-continue feedback. Retrieved from http://www.igd.com/our-expertise/Supply-chain/Logistics/11311/Developing-supply-chain-talent/Developing-your-supply-chain-people/Procter-and-Gamble-360-degree-Stop-start-continue-feedback/

Richardson, J. (2005). Instruments for obtaining student feedback: A review of the literature.

Assessment and Evaluation in Higher Education, 30, 387-415.

Sakui, K. (2004). Wearing two pairs of shoes: Language teaching in Japan. *ELT Journal*, 58, 155-163.

Sato, K. (2002). Practical understandings of communicative language teaching and teacher development. In S. Savignon (Ed.), *Interpreting communicative language teaching* (pp.41-81). New Haven: Yale University Press.

Schön, D. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.

Seldin, P. (1993). The use and abuse of student ratings of professors. *The Chronicle of Higher Education*, 39, 40-43.

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.

Spencer, K., & Schmelkin, L. (2002). Student perspectives on teaching and its evaluation. *Assessment and Evaluation in Higher Education*, 27, 397-409.

Spooren, P., & Van Loon, F. (2012). Who participates (not)? A non-responses analysis on students' evaluation of teaching. *Procededia- Social and behavioral Sciences*, 69, 990-996.

Steinbrecher & Associates, Inc. (1992). Stop/Start/Continue: Tools for Effective Development. Retrieved from www.steinbrecher.com/wp-content/stop-start-continue.pdf.

Stewart, M. (2015). The language of praise and criticism in a student evaluation survey. *Studies in Educational Evaluation*, 45, 1-9.

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.

Svinicki, M. (2001). Encouraging your students to give feedback. In K. Lewis (Ed.), *Techniques and strategies for interpreting student evaluations* (pp.17-24). San Francisco: Jossey-Bass.

Tsurata, Y. (2003). Globalisation and Japanese higher education. In R. Goodman & D. Phillips (Eds.), *Can the Japanese change their education system?* (pp.119-151). Oxford: Symposium Books.

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.

Yao, Y., & Grady, M. (2005). How do faculty make formative use of student evaluation feedback?: A multiple case study. *Journal of Personnel Evaluation in Education*, 18, 107-126.

Appendix 1

This evaluation is for honest and meaningful feedback on your English class. Please take some time and give me, the teacher, some ideas how I can improve the class. Please think about the whole course and not just today's lesson. This evaluation will not affect your grade and should be filled in anonymously. Thank you for your help.

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) _____

Appendix 2

この評価用紙は、授業に対するあなたの正直な考えを問う、私にとっては意義深いものとなります。ご協力お願いします。少し時間を頂戴し、私が今後授業をどのように改善できるか意見を聞かせて下さい。今日の授業だけではなく、この講義全体について振り返ってみて下さい。この評価は、無記名でまたあなたの成績には一切関係ありません。

止めた方がいいこと	現在授業で行っていることで、上手くいっていないことを下記に挙げて下さい。（止めた方がいいこと） 1）＿＿＿＿＿＿＿＿ 2）＿＿＿＿＿＿＿＿ 3）＿＿＿＿＿＿＿＿ 4）＿＿＿＿＿＿＿＿
始めた方がいいこと	授業に導入すると有効ではないかと思うことを下記に挙げて下さい。 1）＿＿＿＿＿＿＿＿ 2）＿＿＿＿＿＿＿＿ 3）＿＿＿＿＿＿＿＿ 4）＿＿＿＿＿＿＿＿
継続した方がいいこと	授業ですでに導入済みで今後も続けていくべきと思うことを下記に挙げて下さい。 1）＿＿＿＿＿＿＿＿ 2）＿＿＿＿＿＿＿＿ 3）＿＿＿＿＿＿＿＿ 4）＿＿＿＿＿＿＿＿