Potential Problems of Student Evaluation of Teaching (SET) in Off-Shore Campuses in Southeast and East Asia and Suggestions

Ulas Basar Gezgin
RMIT University, Vietnam

This paper consists of 3 parts. In the first part, 5 problems of SET are discussed (biases due to SET (which refer to problems due to online surveys, equivalency of SET questionnaires, cognitive biases, and formative vs. summative uses of SETs); biases due to courses; biases due to students; biases due to teachers; and issues with feedback. In Part 2, problems that may be peculiar to off-shore campuses in (South)East Asia are discussed (surface vs. deep learners/teachers; the validity of the questionnaires; curriculum and assessment tasks; and closing the feedback loop). In Part 3, suggestions are listed.

INTRODUCTION

Student evaluations of teaching have a growing significance as they are increasingly used for personnel decisions (Stark-Wroblewski, Ahlering & Brill, 2007). SETs are commonly used for improving courses and teaching, appraisals of the teachers and accountability purposes (Kember, Leung & Kwan, 2002). For some, it is the magic wand that will automatically improve teaching. In the meantime off-shore campuses mushroom in Southeast and East Asia, and SET are widely used in these campuses accompanied with problems due to SET and others due to the nature of the off-shore education services. The first part reviews the research on the problems due to SET, the second part reflects on the problems that may be peculiar to off-shore campuses in Southeast and East Asia, and the final part provides some suggestions.

PROBLEMS OF SET(1)

Worthington (2002) lists 4 possible sources of bias in SET:
1) Factors related to the administration of SET.
2) Factors related to the course such as difficulty and workload.
3) Factors related to students such as age, interest in the course and expected grade.
4) Factors related to the teacher such as experience, appearance, gender etc.

It is not sufficient to have no bias in SET. In the ideal cases where no biases involved, how student feedback is used by teachers makes a difference. If teachers don't use feedback in an appropriate way, a non-biased survey will be of no avail. Thus the use of feedback is considered to be the fifth factor below.

Biases Due to SET

There are 4 biases due to SET. They are; problems due to online surveys, equivalency of SET questionnaires, cognitive biases, and formative vs. summative uses of SETs.

Problems Due to Online Surveys

A plethora of studies focused on the advantages and disadvantages of online surveys. One of the main reasons for a move from classroom surveys to online surveys is the fact that costs of printing, collecting, scanning, administration etc. are reduced by online surveys (Barkhi & Williams, in press; Dommeyer et al., 2004; Watt et al., 2002). Classroom surveys have usually high participation rates compared to web-based surveys (Barkhi & Williams, in press; Dommeyer et al., 2004; Nulty, 2008; Symbaluk & Howell, in press; Watt et al., 2002). Nulty (2008) reviews various studies where the response rate for the latter is 30 odd% less than that of the former. The former is more representative since web-based survey participants have either very negative ('I hate my teacher') or very positive ('I admire my teacher') views (Symbaluk & Howell, in press). Thus, Nulty (2008) discusses the possibility that the online survey respondent characteristics may be different compared to nonrespondent characteristics, which will bias the data. Then the data may not be considered to be representative. This problem is independent of the low response rate. It corresponds to sampling bias. Furthermore, online surveys may be filled by the students who don't attend the classes (Dommeyer et al., 2002).

Dommeyer et al. (2004) and Venette, Sellnow & McIntyre (in press) found no differences in student ratings of instruction with regard to the mode of response (ie online survey vs. traditional classroom survey). However, Venette, Sellnow & McIntyre (in press) found that students provided longer responses to open-ended commentary questions in online mode. It may be due to the fact that students would feel more anonymous in online surveys or that the traditional social barriers are removed in online responding as much as the flaming effect in internet use (Barkhi & Williams, in press). Consistent with this explanation, Barkhi & Williams (in press) found that more extreme comments came out by computer-mediated evaluations compared to classroom surveys confirming Symbaluk & Howell (in press). Diverging from Venette, Sellnow & McIntyre (in press) which found no difference in negativity and positivity of responses on online surveys, Barkhi & Williams (in press) found that on average, the course and the teacher received lower ratings on online surveys compared to classroom surveys, though the effect disappeared when course and teacher variables are statistically controlled. However, Barkhi & Williams (in press) also found that extremely negative evaluations are common on online survey and extremely positive evaluations are common in classroom surveys. Students who are not satisfied by the course are more likely to respond online (Barkhi & Williams, in press).

To boost online response rates, Dommeyer et al. (2004) suggests grade incentives, but they may not be feasible since it will take more time to process extra marks. Nulty (2008) lists the three most common ways to increase response rates for online surveys:

> "(1) repeat reminder emails to non-respondents (students); (2) repeat reminder emails to survey owners (academics); (3) incentives to students in the form of prizes for respondents awarded through a lottery" (Nulty, 2008, p.303).

Nulty (2008) makes a list of other suggestions that emerge out of the review of previous research. Though the majority of these suggestions are implemented by default in the current settings such as brief questionnaires rather than long ones, and anonymity, one out of these is significant: "Persuade respondents that their responses will be used" (Nulty, 2008, p.305). However, this suggestion applies to any mode of survey. The students should be convinced that "their opinions do matter" (Spencer & Schmelkin, 2002, p.406) to boost the participation to SETs. Chen & Hoshower (2003) finds that contributing to the improvement of teaching and course are the main motivators for students' participation to SET. Students are more motivated if they are convinced that their responses will be used for improvement (Chen & Hoshower, 2003). Thus rather than behavioral motivators such as extra marks or prizes, cognitive motivators should be employed to make SET meaningful for students.

Equivalency of SET Questionnaires

Various SET questionnaires are used in different countries and in different universities in a single country. Some questionnaires include 5 items whereas others include more than 30 items.[(2)] Short questionnaires take less time, but they are less valid as the phenomena they are supposed to measure are

inherently multidimensional. SET questionnaires are used to measure teaching quality or teaching effectiveness, but they may tap different dimensions. As a result, one should be cautious in interpreting the findings discussed in this paper. A comparative study is necessary where different SET questionnaires are administered to a local student population to check the consistency of the results.

Cognitive Biases

According to Barkhi & Williams (in press), the following cognitive biases may be involved in student ratings: Hindsight bias, confirmation bias, self-serving bias, recency bias and serial position bias. It seems that the latter three are the most relevant ones for a study of student ratings: Students attribute success to themselves and failure to teachers and other external factors (self-serving bias); since the recent events are more salient in human memory, students may be biased towards recall of recent weeks of the semester (recency bias) and the responses are affected by the order through which the questions are asked (serial position bias). Yorke (in press) adds 'acquiescence bias' or 'yea-saying' which involves the tendency of the respondents towards affirming answers. Yorke (in press) also mentions 'primacy effect' as positive items such as 'strongly agree' runs from left to right which means they are always read earlier than negative items. However Yorke (in press) could not find any evidence for ordering effect and acquiescence bias in his empirical study. Thirdly, 'indifference bias' refers to respondents' tendency to choose the mid value (i.e. 3 over 5-point Likert scale) or only one value in a uniform manner (i.e. 2 for all questions) (Yorke, in press). Indifference bias in the second sense is assumed to be eliminated by the use of negative items along with positive items, though it is not always eliminated.

As a further potential cognitive bias, Yorke (in press) discusses the mid-point in Likert scales. It may have three senses for respondents: It may refer to their true neutrality or ambivalence or non-applicability. Alternatively, forced choice items can be used (ie 4 items by removing the mid-point) but then this may cause a higher level of error since the respondents are not provided with the opportunity to be neutral or ambivalent (Yorke, in press).

More studies need to be conducted on these cognitive biases.

Formative vs. Summative Uses of SETs

One problem that stains student ratings is the fact that the evaluation process is summative, i.e. it is at the end of the process. Barkhi & Williams (in press) suggests that by utilizing the ease of administration of online surveys, the feedback process can be made continuous throughout the semester so that the teachers can improve their teaching quality based on more frequent feedback.

Nulty (2008) states that a response by even a single respondent is relevant to improve the teaching quality; however, teaching evaluation surveys are increasingly used for summative purposes and for performance evaluation which is untenable as the sampling requirements are not met in many of the surveys.

Since SETs are usually done at the end of the semester and the end of the semester is too late to improve the course based on student feedback, Spencer & Schmelkin (2002) suggests that SETs need to be complemented by mid-semester student feedback to boost the teaching quality.

To sum up, the disadvantages of online surveys, the equivalency of the SET questionnaires, cognitive biases and formative vs. summative uses of SET data should be addressed to obtain more valid and reliable SET results.

Biases Due to Courses

Using RateMyProfessor.com (RMP) data, Felton et al. (2008) finds a positive relationship between course easiness and teacher's attractiveness, and course quality ratings. The second link demonstrates that halo effect operates in course evaluations. If students think that a teacher is attractive and course is easy, they rate the course as a high quality course (Felton et al., 2008). Bowling (2008) finds that easiness of the course is associated with favorable ratings in course experience surveys in low-ranked schools, while the positive correlation is relatively lower for high-ranked ones. In both schools, the same correlation is observed at different levels. In other words, the easier the course, more favorable are the ratings. Bowling

(2008) attributes this difference to the fact that students at high-ranked schools are more motivated to learn more, and thus easiness is not a very significant criterion for their evaluations of a course or a teacher. Likewise, Davison & Price (2009) and Felton, Mitchell & Stinson (2004) found that high ratings are positively related to easiness on RMP data. In other words, professors teaching easy courses with low course load are rated highly positively (Davison & Rice, 2009).

A simple measure of easiness is the failure rates. In evaluating the teaching performance, failure rates should be considered to adjust the high ratings for easy courses and low ratings for difficult courses.

Biases Due to Students

Senior and junior students rate the courses and teachers more positively compared to sophomores and freshman students (Barkhi & Williams, in press). Students with higher GPAs tend to rate their teachers more favorably (Betoret & Tomas, 2003; Chen & Hoshower, 2003). Worthington (2002) found that students who expect a higher grade rated their teacher more favorably. Finally, students are more likely to rate their professors when they want to complain about them (Davison & Rice, 2009). To minimize the biases due to students, SET should include questions asking expected grade and current GPA. These questions are already part of some SETs (eg Worthington, 2002).

Biases Due to Teachers

Felton, Mitchell & Stinson (2004) finds that teachers who are considered to be attractive are rated more favorably. Though this finding is based on RMP data, it casts doubt on reliability of SETs as they may be subject to similar halo effects. Based on this finding, it may be inferred that elderly teachers are more likely to be rated less favorably.

Using RMP data, Symbaluk & Howell (in press) found that Teaching-Award winners receive higher ratings compared to Research-Award winners, although the latter still receives positive ratings. While this result may be trivially true, it points out a trade-off between teaching orientation and research orientation that is common in many higher education settings.

Ogier (2005) finds that teachers who are non-native speakers of English receive lower ratings for overall satisfaction with the course. This may lead to a halo effect for teachers who are native speakers of English. Furthermore, ESL students i.e. students coming from non-English speaking backgrounds (non-ESB) rate non-ESB teachers far lower compared to ESB students (Ogier, 2005). Finally, non-ESB students are found to rate their teachers more favorably in general (Worthington, 2002).

Shevlin et al. (2000) finds that SET responses are largely affected by teacher's personal attributes such as charisma. Furthermore;

> "[t]here is a high correlation between the ratings produced by students taking different course units taught by the same teacher, but little or no relationship between the ratings given by students taking the same course unit taught by different teachers. This suggests that students' evaluations are a function of the person teaching the course unit rather than the particular unit being taught" (Richardson, 2005, p.389).

Whilst student evaluations provide a significant source of feedback for the university, it may have detrimental consequences. To boost the ratings, the teachers may be inclined towards lenient marking, as high grades will likely lead to higher student ratings (Stark-Wroblewski, Ahlering & Brill, 2007). Even worse may come out:

> "If, for example, the student ratings reflect instructor charisma, sense of humor, grading leniency or entertainment value rather than learning, then instructors will be motivated to achieve something not in the interest of higher education" (Otto et al., 2008, p.356).

Teachers who receive high ratings see SET as highly valid (Nasser & Fresko, 2002). "Instructors support course evaluation when it flatters them and oppose it when they are criticised" (Nasser & Fresko, 2002, p.197). Thus SET reproduces itself even if it may not be efficient. This point involves teachers' self-serving biases. They may falsely think that mastery of the subject guarantees quality teaching. In the

lines of the attribution theories, just as psychologically healthy students attribute their failure to external and unstable factors such as effort and teachers; teachers may attribute students' failure to lazy or 'dumb' students, although in some cases, students' failure may be due to teachers' failure.

Stark-Wroblewski, Ahlering & Brill (2007) confirmed that SET scores were not a measure of learning outcomes. In other words, high scores on SET do not show whether students learned the topics or not. Learning outcomes explained the variances in grades more than SET ratings. Thus, Stark-Wroblewski, Ahlering & Brill (2007) suggests the use of learning measures along with SET in evaluating teaching effectiveness.

Issues with Feedback

Improvement in quality of teaching depends on how teachers receive, interpret and utilize SET responses (Ballantyne, Borthwick & Packer, 2000). Although students are not the only stakeholders in SET, the literature is flooded by research on students' perceptions and characteristics only. One exception is Shao, Anderson & Newsome (2007) in which a large sample of university administrators (deans and heads) are asked to rate the current state of 'evaluating teaching effectiveness'. In another exception, Hendry, Lyon & Henderson-Smart (2007) investigated teachers' perceptions of SET, whether they used them to boost teaching quality, and in what ways they did this, such as improving the slides, decreasing the number of topics covered in the course, reducing the theoretical parts, encouraging a higher level of participation etc.[3]

Seeing the drop of SET ratings for 3 departments throughout years, Kember, Leung & Kwan (2002) questions the quality of teaching in their university. They provide 5 possible explanations for this downward trend that are not mutually exclusive:

- A stable plateau is reached in the ratings. Kember, Leung & Kwan (2002) thinks that evidence for this explanation is insufficient.
- The use of feedback data is ineffective. This is a likely possibility.
- Incentives to use feedback data are lacking. This is another likely possibility. The majority of teachers feel that good teaching is not sufficiently rewarded.
- Appraisal function of SETs crowds out its formative function. SET is considered to be an instrument for contracts and tenure rather than improving courses.
- Metric problems with SET lead to instability in ratings.

The last 4 points provide ways to improve SET results, though they don't automatically guarantee an improvement in teaching quality: Refined ways of collecting quality feedback need to be implemented; incentives should be used to encourage the use of student feedback by teachers; formative and summative functions of SET should be addressed in two different questionnaires if possible; and SETs should be revised for higher levels of validity and reliability.

PROBLEMS THAT MAY BE PECULIAR TO OFF-SHORE CAMPUSES IN EAST AND SOUTHEAST ASIA

Four major problems are observed in off-shore campuses in East and Southeast Asia. These are the problem of surface learners/teachers vs. deep learners/teachers, the validity of the questionnaires, curriculum and assessment tasks, and closing the feedback loop.

Surface Learners/Teachers vs. Deep Learners/Teachers

A neglected factor in student evaluations is students' expectations for the course (Fisher & Miller, 2007). These expectations are more or less forged by education prior to tertiary level. Leung (2001) identifies 6 dichotomies that characterize mathematics education in East Asia, to some extent:

- 'Product vs. process': East Asian focus is the product of problem solving rather than the process of solving it.
- 'Rote learning vs. meaningful learning': The former is dominant over East Asia.

- 'Studying hard vs. pleasurable learning': Studying hard is highly appreciated in East Asia.
- 'Extrinsic vs. intrinsic motivations': East Asian students are more extrinsically motivated e.g. by exams.
- 'Whole class teaching vs. individualized learning': Whole class teaching is preferred in East Asia.
- 'Subject matter vs. Pedagogy': Mastery of the subject is considered to be sufficient to teach a topic in East Asian context following the traditional transmission model of teaching whereby teachers as authorities transmit their knowledge; and fill students' minds with their knowledge.

Almost all the students of off-shore campuses in East and Southeast Asia are Asian. As a result, they may come with different expectations of education which would be in clash with the teaching model underlying international universities. Following the dichotomies above, Asian students may expect to receive education where product, rote learning, studying hard, extrinsic motivation, whole class teaching and teacher as a high authority figure are valued which are squarely opposite to the notion of international education. A teacher who prefer the second pole of the dichotomies (ie process, meaningful learning, intrinsic motivation, individualized learning, and pedagogy)[4] may be rated lower on SET. In this vein, two points on Macalpine (2001)'s list of problems with use of SET are particularly applicable to off-shore campuses in Asia:

"- Some students may dislike being required to think rather than memorise by rote;
- Some students may dislike being expected to answer questions orally (in their second language), in class, rather than sitting passively" (Macalpine, 2001, p.565).

Supposing that Asian students are harboring the first poles of the six dichotomies, the teacher and course which would get the highest ratings on SET would have the following properties:

- Guaranteeing passing marks by rote learning.
- Whole class teaching rather than asking questions individually.
- No innovative teaching as it would individualize the process.
- Multiple choice tests rather than essay-type exams.
- Written exams rather than oral exams.
- Sample questions to prepare for the final exam.

These characterize surface learners as opposed to deep learners. Surface learners which are common in East Asia may pull surface teachers as a magnet. Both deep learners and deep teachers would suffer. Surface learners will rate surface teachers more favorably within the current SET system. In this context, Lyon & Hendry (2002) finds that Course Experience Survey widely used in Australia is based on a traditional education model and does not fit a problem-based education model. Thus SETs should be revised to embrace innovative education models.

Keeping in mind that SET results may be biased due to different conceptions of learning and effective teaching held by students (Hendry, Lyon & Henderson-Smart, 2007), SET questionnaires should be revised to detect these conceptions. To identify students' expectations, a formative expectations questionnaire is necessary. Secondly, as the use of multiple choice tests that are known to discourage deep learning (Struyven, Dochy & Janssens, 2005) furthers surface learning in international universities of Asia, essay type exams and alternative assessment methods that are known to encourage deep learning (Struyven, Dochy & Janssens, 2005) need to be promoted.[5] Finally, SET system should be clear on how rewarding of surface teaching/learning would be avoided. SET questionnaires which are based on the transmission model of teaching should be revised to cover alternative teaching methods.

On the other hand, Ballantyne, Borthwick & Packer (2000) finds that in general, teachers focus more on improving student learning, whereas students focus more on assessment. Therefore, the distinction of surface learner/teacher vs. deep learner/teacher and other issues discussed in this section may not be specific to Asia. More studies are necessary to compare expectations of Asian and non-Asian students.

The Validity of the Questionnaires

SET questionnaires are always validated in the central hub, but then they are used in off-shore campuses as well. They may not be valid for off-shore campus student bodies. Thus, formal validation of the questionnaires in peripheral hubs is necessary.[6]

A second issue related to the first one would be the selection of SET questionnaire(s) for use. An option is having more than one SET questionnaire that would be valid in different campuses. A second option is using a single SET questionnaire that would be validated in all campuses. A third option is using a single questionnaire validated in all campuses along with a second SET questionnaire that will serve the local educational issues and needs. By this option, extra questions can be added for off-shore campuses such as satisfaction with the coordination across the off-shore campus and the central hub.

A third issue will be the language of SET. The reception of the SET questionnaire by off-shore campus students would be different, since it is in their second language. It is more likely to receive genuine feedback in their native language rather than English. Thus, off-shore campus students may be allowed to answer SET questionnaire in their native language. Although this will increase the administrative costs, it will worth the effort, as this may be another reason for low response rates on online surveys in offshore campuses.

Curriculum and Assessment Tasks

Assessment tasks prepared in the central hub may not fit the offshore campus. For one thing, the questions may be ethnocentric. Paralleling the issue of the validity of SET questionnaires across campuses, three options emerge:

1) Using all the materials and assessment tasks used in the central hub without considering any local feasibility problems.
2) Using materials and assessment tasks that are developed in various campuses.[7]
3) Using both central hub's national materials and assessment tasks along with materials and assessment tasks that will fit the local needs of the off-shore campus.

Where these problems are not addressed, student satisfaction will be reflected as low ratings for the teacher who will be a scapegoat in that case. Students would provide lower ratings not because the course has low quality, but the rapport between the central hub and offshore campuses is asymmetric or central hub is not sensitive to the local needs. These problems are intermingled with the issue of standardization whereby the off-shore campus students receive identical diplomas with central hub students.

Closing the Feedback Loop

Although it is easy to see SET's invasion of tertiary education as the injection of the Costumer Satisfaction Model into universities, the comparison is misleading. Customers will not be satisfied with a low-quality product. If the product does not come out as it is advertised, this won't make the customers happy. However, students may be satisfied with low-quality education as they may prefer to study less and guarantee a passing mark without effort. This problem is aggravating in off-shore campuses since the teachers may be a toast between the students and the central hub's administrative structure. Spencer & Schmelkin (2002) suggests that teachers' and administrators' evaluation of SET process should be carried to the equation for a further improvement of teaching. In the off-shore context, SET should be accompanied with teachers' and administrators' evaluation of teaching/learning, central hub coordination and off-shore administration. By covering all stakeholders, more genuine feedback will be collected.

SUGGESTIONS

Shao, Anderson & Newsome (2007) lists the following items that can be used for evaluating teaching effectiveness: Current in field, student evaluation scores, student written comments, Chair's evaluation, Teaching Awards, peer's evaluations, intellectual contributions, Teaching Portfolio, classroom visits, Dean's evaluation, class assignments, use of technology, alumni feedback, colleagues' opinions, grade distribution, course notebooks, course level (grad/undergrad), course type (required/elective), class

enrolment, and drop rate (Shao, Anderson & Newsome, 2007, p.361). None of them will be reliable to evaluate teaching performance by their own. A teacher may hold current knowledge in field, but this does not mean s/he has an appropriate pedagogical approach. Student evaluation scores and student written comments may be biased as discussed in previous parts. Dean's and Chair's evaluation may be misleading, as s/he may not have direct experience of the teacher's performance. The range of Teaching Awards is quite narrow as the recipients are just a few, and the jury may be subjective if clear guidelines and transparency are not the norm. Peer's evaluation may be subjective as much as self-ratings. Intellectual contributions are difficult to measure. Classroom visits are very useful; however they may not be representative of the teaching performance, compared to all the classes throughout the semester.[8] Class assignments, course notebooks, course level and course type, class enrolment and drop rate will not provide much information without the use of other methods. Finally, an evaluation of use of technology may lead to technology fetishism where technology would be deified at the expense of knowledge and skills. To weather the deficiencies of SETs, Macalpine (2001) suggests a combination of SETs, in-class peer evaluations and teaching portfolios. He states that deficiencies of the three will be balanced by the use of them together.

A major avenue to boost teaching quality is to improve SET methods. In this vein, Huxham et al. (2008) compares the results of student ratings based on questionnaire with ratings by other methods such as focus groups, student diaries, and rapid feedback.[9] Whilst the qualitative responses to course experience questionnaire highly overlapped with those by other methods, students considered the latter as time-consuming and too demanding. However, focus groups were favored for a deeper level of reflection on the course. Furthermore, some students also questioned the reliability of the course experience surveys (Huxham et al., 2008). Huxham et al. (2008) stresses the significance of qualitative responses for formative evaluation i.e. to improve teaching, and that of quantitative responses for summative evaluation i.e. for managerial decisions.

In this paper, 5 problems of SET are discussed. These are

1) Biases due to SET which refer to
a) Problems due to online surveys
b) Equivalency of SET questionnaires
c) Cognitive biases
d) Formative vs. summative uses of SETs
2) Biases Due to Courses
3) Biases Due to Students
4) Biases Due to Teachers
5) Issues with Feedback

For Problem 1(a), incentives to boost response rates are necessary. In cases where they don't work, a cautious interpretation of the results is a must. The results should be considered with regard to the limitations imposed by low response rates (Barkhi & Williams, in press).[10]

For Problem 1(b), more studies should be conducted to ensure that different questionnaires of SET tap the same dimensions of teaching performance.

For Problem 1(c), more research is needed as the relevant studies are a scarcity.

For Problem 1(d), the formative and summative uses of SETs should be demarcated. To serve this purpose, two different SET questionnaires may be used. The formative questionnaire may cover qualitative responses while summative questionnaire may cover both qualitative and quantitative responses to double-check each other, following the suggestion by Huxham et al. (2008).

For Problem 2, course easiness should be checked. This can be checked by failure rates. A simple formula can be used to adjust the high ratings of easy courses and low ratings of difficult courses based on failure rates.

For Problem 3, expected grade and current GPA can be asked in SET questionnaires.

Due to Problem 4, the following issues should be kept in mind while evaluating teaching performances:

- Teachers who are considered to be less attractive are rated less favorably.

- Elderly teachers are rated less favorably.
- Teachers who are not native speakers of English are rated less favorably.

Finally, the university should take measures against the abuse of SET where teaching quality is compromised to please the students in a move towards a boost of SET results.

Lastly, to address Problem 5, incentives should be provided for teachers so that they can use the feedback to improve the teaching quality.[11] Two papers worth mention in this regard are the following: Chen & Hoshower (2003) suggests that course syllabus may contain brief information about how SET results are used to improve the course. This will increase the perceived value of SET through the eyes of students as well as teachers (Chen & Hoshower, 2003). Secondly, Ballantyne, Borthwick & Packer (2000) reports the development of booklets on main issues such as interactive teaching strategies, teaching for learning, evaluating the quality of teaching etc. based on SET responses and teachers' perceptions. This simple and brilliant idea can be applied in off-shore campuses as well.

In Part 2, problems that may be peculiar to off-shore campuses in East and Southeast Asia are discussed. These are:

1) Surface Learners/Teachers vs. Deep Learners/Teachers
2) The Validity of the Questionnaires
3) Curriculum and Assessment Tasks
4) Closing the Feedback Loop

As to Problem 1, students' perceptions and teachers' perceptions may be different as shown by Ballantyne, Borthwick & Packer (2000). Therefore expectations of students and teachers should be clarified on the first week. The university should ensure that surface teachers will not be rewarded due to high SET results that do not reflect the teaching quality.

For Problem 2, cross-validation studies are necessary, whereas parallel decisions need to be taken for the cross-campus use of curriculum and assessment tasks developed in the central hub to manage Problem 3.

As to Problem 4, to close the loop, Betoret & Tomas (2003) lets teachers to evaluate their students by parallel items such as 'the students are (the teacher is) motivated' etc. Not only teachers, but all the stakeholders of off-shore education services should be part of the evaluation loop to improve the teaching quality.[12]

This paper makes practical suggestions whilst proposing new research topics for researchers. These suggestions should be tested by practice, and these tests will definitely lead to newer research topics.

ENDNOTES

(1) In this paper the terms 'evaluation' and 'rating' are used interchangeably for practical purposes although "[a]n 'evaluation' implies a conclusion based on some direct definitive measure, whereas a 'rating' connotes data susceptible to interpretation" (Otto et al., 2008, p.365).

(2) For examples of various SET questionnaires, one can check Barkhi & Williams, in press; Cohen, 2005; Kember, Leung & Kwan, 2002; Macalpine, 2001; Neumann, 2000; Ogier, 2005; Shevlin et al., 2000; Venette, Sellnow & McIntyre, in press; and Worthington, 2002.

(3) A third exception is Ballantyne, Borthwick & Packer (2000) where parallel items are used for students and teachers.

(4) The dichotomy of studying hard vs. pleasurable learning does not fit the situation well. It has to be elaborated further.

(5) Multiple choice tests are preferred since they are less time-consuming to administer and mark. However even a superficial cost-benefit analysis will show that the pedagogical costs of them weigh heavier than their benefits.

(6) An example is Byrne & Flood (2003) where CEQ (an Australian questionnaire) is validated for Ireland.

(7) This requires the switch of central hub to these internationalized materials and assessment tasks, rather than following its national materials and assessment tasks.

(8) Macalpine (2001) discusses the problems with class observations.

(9) Rapid feedback refers to 3 general open-ended questions to be completed in 5 minutes.

(10) What level of response rate would be sufficient for the validity of the student ratings? Nulty (2008) provides a very useful table to answer this question –assuming that the respondents are randomly selected. However, convenience sampling rather than random sampling is used in course evaluation surveys (Nulty, 2008). Nevertheless, the table is useful since it sets the higher standard for the response rates. For instance, by a liberal statistical standard, 14 (48%) responses are required out of a class of 30 students while 29 responses (96%) are necessary by a stringent standard. As the class size increases, required response rates decrease. E.g. for a class size of 2000, which is highly unlikely, 1% response rate is plausible under a liberal standard.
(11) The distinction between extrinsic and intrinsic motivation applies to teachers as well.
(12) However, it should be kept in mind that SET poses ethical challenges, as the managerial access to and the publicization of the results involve questions on confidentiality and violation of privacy (McCormack, 2005). McCormack (2005) lists the following ethical values relevant to SET: Autonomy and justice, respect for students' opinions, anonymity and confidentiality, and privacy.

REFERENCES

Ballantyne, R., Borthwick, J. & Packer, J. (2000). Beyond student evaluation of teaching: identifying and addressing academic staff development needs. *Assessment & Evaluation in Higher Education*, 25, (3), 221-236.

Barkhi, R. & Williams, P. (in press). The impact of electronic media on faculty evaluations. *Assessment & Evaluation in Higher Education*.

Betoret, F.D. & Tomas, A.D. (2003). Evaluation of the university teaching/learning process for the improvement of quality in higher education. *Assessment & Evaluation in Higher Education*, 28, (2), 165-178.

Bowling, N.A. (2008). Does the relationship between student ratings of course easiness and course quality vary across schools? The role of school academic rankings. *Assessment & Evaluation in Higher Education*, 33, (4), 455-464.

Byrne, M. & Flood, B. (2003). Assessing the teaching quality of accounting programmes: an evaluation of the Course Experience Questionnaire. *Assessment & Evaluation in Higher Education*, 28, (2), 135-145.

Chen, Y. & Hoshower, L.B. (2003). Student evaluation of teaching effectiveness: an assessment of student perception and motivation. *Assessment & Evaluation in Higher Education*, 28, (1), 71-88.

Cohen, E.H. (2005). Student evaluations of course and teacher: factor analysis and SSA approaches. *Assessment & Evaluation in Higher Education*, 30, (2), 123-136.

Dommeyer, C.J., Baum, P., Chapman, K.S. & Hanna, R.W. (2002). Attitudes of business faculty towards two methods of collecting teaching evaluations: Paper vs. online. *Assessment & Evaluation in Higher Education*, 27, (5), 455-462.

Dommeyer, C.J., Baum, P., Hanna, R.W. & Chapman, K.S. (2004). Gathering faculty teaching evaluations by in-class and online surveys: their effects on response rates and evaluations. *Assessment & Evaluation in Higher Education*, 29, (5), 611-623.

Felton, J., Koper, P.T., Mitchell, J. & Stinson, M. (2008). Attractiveness, easiness and other issues: student evaluations of professors on Ratemyprofessors.com. *Assessment & Evaluation in Higher Education*, 33, (1), 45-61.

Felton, J., Mitchell, J. & Stinson, M. (2004). Web-based student evaluations of professors: the relations between perceived quality, easiness and sexiness. *Assessment & Evaluation in Higher Education*, 29, (1), 91-108.

Fisher, M. & Miller, D. (2007). Responding to student expectations: a partnership approach to course evaluation. *Assessment & Evaluation in Higher Education*, 33, (2), 191-202.

Hendry, G., Lyon, P.M. & Henderson-Smart, C. (2007). Teachers' approaches to teaching and responses to student evaluation in a problem-based medical program. *Assessment & Evaluation in Higher Education*, 32, (2), 143-157.

Huxham, M., Laybourn, P., Cairncross, S., Gray, M., Brown, N., Goldfinch, J. & Earl, S. (2008). Collecting student feedback: a comparison of questionnaire and other methods. *Assessment & Evaluation in Higher Education*, 33, (6), 675-686.

Kember, D., Leung, D.Y.P. & Kwan, K.P. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment & Evaluation in Higher Education*, 27, (5), 411-425.

Leung, K.S. (2001). In search of an East Asian identity in mathematics education. *Educational Studies in Mathematics*, 47, 35-51.

Lyon, P.M. & Hendry, G.D. (2002). The use of the Course Experience Questionnaire as a monitoring evaluation tool in a problem-based medical programme. *Assessment & Evaluation in Higher Education*, 27, (4), 340-352.

Macalpine, M. (2001). An attempt to evaluate teaching quality: one department's story. *Assessment & Evaluation in Higher Education*, 26, (6), 563-578.

McCormack, C. (2005). Reconceptualizing student evaluation of teaching: an ethical framework for changing times. *Assessment & Evaluation in Higher Education*, 30, (5), 463-476.

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

Neumann, R. (2000). Communicating Student Evaluation of Teaching results: Rating interpretation guides (RIGs). *Assessment & Evaluation in Higher Education*, 25, (2), 121-134.

Nulty, D.D. (2008). The adequacy of response rates to online and paper surveys: what can be done? *Assessment & Evaluation in Higher Education*, 33, (3), 301-314.

Ogier, J. (2005). Evaluating the effect of a lecturer's language background on a student rating of teaching form. *Assessment & Evaluation in Higher Education*, 30, (5), 477-488.

Otto, J., Sanford, Jr, D.A. & Ross, D.N. (2008). Does my ratemyprofessor.com really rate my professor? *Assessment & Evaluation in Higher Education*, 33, (4), 355-368.

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

Shao, L.P., Anderson, L.P. & Newsome, M. (2007). Evaluating teaching effectiveness: where we are and where we should be. *Assessment & Evaluation in Higher Education*, 32, (3), 355-371.

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

Spencer, K.J. & Schmelkin, P. (2002). Student perspectives on teaching and its evaluation. *Assessment & Evaluation in Higher Education*, 27, (5), 397-409.

Stark-Wroblewski, K., Ahlering, R.F. & Brill, F.M. (2007). Toward a more comprehensive approach to evaluating teaching effectiveness: Supplementing student evaluations of teaching with pre-pist learning measures. *Assessment & Evaluation in Higher Education*, 32, (4), 403-415.

Struyven, K., Dochy, F. & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: a review. *Assessment & Evaluation in Higher Education*, 30, (4), 325-341.

Symbaluk, D.G. (in press). Web-based student feedback: comparing teaching-award and research-award recipients. *Assessment & Evaluation in Higher Education*.

Venette, S., Sellnow, D. & McIntyre, K. (in press). Charting new territory: assessing the online frontier of student ratings of instruction. *Assessment & Evaluation in Higher Education*.

Watt, S., Simpson, C., McKillop, C. & Nunn, V. (2002). Electronic course surveys: does automating feedback and reporting give better results? *Assessment & Evaluation in Higher Education*, 27, (4), 325-337.

Worthington, A.C. (2002). The impact of student perceptions and characteristics on teaching evaluations: a case study in finance education. *Assessment & Evaluation in Higher Education*, 27, (1), 49-64.

Yorke, M. (in press). 'Student experience' surveys: some methodological considerations and an empirical investigation. *Assessment & Evaluation in Higher Education*.