Student Evaluation of Teaching surveys: do students provide accurate and reliable information?

Tek Lama  
College of Business  
Victoria University  

tek.lama@vu.edu.au

Patrick Arias  
College of Business  
Victoria University  

Krista Mendoza  
College of Business  
Victoria University  

Joanne Manahan  
College of Business  
Victoria University  

Abstract

Purpose- This paper explores patterns of students' response behaviour of international students studying in an Australian university when filling out student surveys evaluating lecturers and courses. The study focuses on whether information obtained through the survey process can be relied upon to make management decisions.

Design/methodology/approach- Data for this study was collected from the Student Evaluation of Teaching and Student Evaluation of Units surveys of 74 students across four different classes at a university campus in Australia which caters education solely to international students. The data was analysed using both descriptive and econometric analytical methods.

Results / findings- The results of the study seem to suggest a reasonable level of diligence is lacking on the students' part in answering the surveys, raising a concern about the reliability of information. This tendency seems to be prevalent among all students irrespective of their gender and nationality.

Implications- This study raises issues with respect to the execution process of student feedback surveys and stresses need for a careful consideration of the use of the survey information in making various management decisions.

Key words: Student survey, Student evaluation of teaching, Student evaluation of units, student feedback.

JEL Classification: I21  
PsycINFO Classification: 3560  
FoR Code: 1503  
ERA Journal ID#: 123340
Introduction

It is a common practice in Australian universities to conduct a Student Evaluation of Teaching survey and using the student feedback as an instrument for assessing lecturers’ performance. Most university lecturers in Australia are evaluated using this process every year. The principle underpinning the survey is that the feedback enables universities to identify areas needing improvement, develop new teaching strategies, improve delivery quality and boost teaching outcomes. While the aim is noble, it is debatable as to whether the process used is generating reliable information consistent with its intended objective.

Blower (2014) pointed out the severe implications of such a survey process on academics. He reported that some lost their jobs, while others are considering leaving the profession altogether, primarily because they feel they are subjected to ‘continuous undermining and criticism’ mostly for pretty lame reasons. There is even a suggestion that a number of high calibre academics receive criticism on minor issues such as their taste in clothing, accent (i.e. pitch in pronunciation) and style of delivery.

The concern here is not about academics receiving negative assessments or criticism. There are bound to be some who are not meeting required expectations. It is imperative that such individuals are identified so university management can initiate appropriate measures to address the issue. This is what the Student Evaluation of Teaching survey process is designed to achieve. The concern however, is whether students give a considered response that contains accurate information about academics which can be reliably used for decision making purposes.

While the issues highlighted with respect to student survey process may exist irrespective of institutions (university or private higher education provider) or type of students (local or international or mix), this study particularly focuses on a university which provides education exclusively to the international students. Two reasons underpin this choice. First, universities across Australia use a common set (if not identical) of student survey process which private higher education providers may not necessarily follow. This provides better representation. Second, international students differ from their local Australian counterparts with respect to their language, culture, system of education, and motivation and; we argue that this provides interesting context to examine the issues.

It is argued that, when answering anonymous surveys, respondents tend to randomly tick the boxes without understanding the questions being asked. Such behaviour is likely to occur for two reasons. First, undergraduate students have not quite reached the stage where they can be expected to think maturely and have a sense of obligation with respect to doing the right thing. Nor are they likely to have the capacity to see the potential consequences on the person concerned of their ticked responses. Second, they are evaluating someone who, most likely, has been giving them perceived criticism about their work, and students may feel that this is their payback time.

Inaccurate and/or biased information will lead to decisions with potentially negative consequences. This is also the case with Student Evaluation of Teaching survey. An unfair treatment of academics based on ill-considered and possibly biased responses can significantly damage both the academic and the university concerned. Thus, the purpose of this study is to explore the survey process. The survey process involves students having to fill out both Student Evaluation of Teaching and Student Evaluation of Units survey questionnaires towards at the end of each semester at the absence of the lecture being evaluated. In this study, it is argued that Student Evaluation of Teaching survey data lack accuracy and reliability and this argument is supported with empirical evidence.

Literature Review

The issue of whether student feedback is considered valid and reliable has been debated for years, both in Australia and in other countries. Studies have questioned student feedback as useful based on the psychological implications of students assessing their lecturers and the statistical validity and reliability of the instrument (Slade & McConville, 2006). For
example, the motivations and expectations of students with respect to their grades could bias their lecturer ratings (Cashin, 1988). In other words, students are likely to rate a lecturer who shows a greater level of leniency more highly; and this may lead some lecturers even to compromise their academic integrity in awarding marks. This behaviour is consistent with human rationality.

Tomasco (1980) stated that students may evaluate a lecturer on personality rather than on teaching effectiveness and ability. In most cases, individuals’ perceived personality and charm speak louder than their talent or effective communication. In other words, students may be choosing form over substance, resulting in incorrect assessment of the academics concerned.

Other researchers, however, term such criticisms as ‘myths’, usually created by those in the faculty being judged negatively in surveys. Aleamoni (1999) found that 16 common student feedback myths were simply myths, but that student rating results were nevertheless being misinterpreted and misused.

Applying the notion that teaching is multidimensional, Marsh and Roche (1999) went on to prove the usefulness of Student Evaluation of Teaching survey. They argued that the only way for such surveys to retain its validity and usefulness is if it manages to reflect, in its distinct components, the multidimensionality of effective teaching. If this is the case, the Student Evaluation of Teaching surveys are a function of the instructor who teaches a course rather than the course that is taught. Accordingly, it should be relatively unaffected by the various potential biases.

Richardson (2005) concluded from a review of the literature that student feedback surveys are generally well established in the United States and in Australia. He further added that it has become more popular in the United Kingdom. The results of pilot surveys he conducted serve to support the claim that student feedback is an important tool for assessing the quality of both teachers and the programme. However, he highlighted two issues concerning the cognitive process behind student responses: the possibility of students misinterpreting and the likelihood of their lacking a sense of responsibility.

The studies which have made the greatest contribution to the study of Student Evaluation of Teaching are Ferguson (2010) and Balch (2012). Ferguson explains the origin of the Tripod Survey and its underlying features which have been used in elementary and secondary schools across the United States, Canada and China over the past decade. Ferguson also designed a test to be conducted in large secondary school classrooms that were chosen to take part in the Method of Effective Teaching (MET) project. His work has been revised and improved over and over again in further studies and tests. Balch’s (2012) dissertation, on the other hand, was aimed at developing and validating a student survey on teacher practice. The survey was constructed using a validity framework, the ‘construct validity.’ It was substantiated by establishing content validity, convergent validity, and predictive validity. Results show there was a positive relationship between value-added student achievement, academic-student engagement, and academic self-efficacy. There was also a strong correlation between a teacher’s score and value-added rankings in social studies and in science.

In addition, a synthesis of existing research was commissioned by the National Comprehensive Center for Teaching Quality (NCCTQ) with the primary focus of providing a better understanding of what constitutes effective teaching and the advantages and disadvantages of a range of measurement tools (Goe, Bell, & Little, 2008). One of the tools discussed was the student rating which was found to be valid and that it provides valuable information as an outcome of students’ interaction with the lecturer. They argued that the student survey is a viable addition to the arsenal of tools in measuring academic effectiveness. But other studies such as Follman (1992) and Worrell and Kuterbach (2001) have asserted that the student survey is not a stand-alone measurement tool because students do not have adequate knowledge of certain criteria associated with teacher effectiveness. Instead they should be used in conjunction with other tools to assess teachers and course programmes.

From the literature, it can be reasonably concluded that Student Evaluation of Teaching survey is a valid tool for measuring the effectiveness of teaching against the backdrop
of criticisms of bias — but only to a limited extent. It is also apparent that no studies have been identified that explore the possibility of students answering the questionnaire in a haphazard manner, randomly ticking boxes without thoroughly reading each question and understanding its weight and meaning. Given that the results of such surveys affect staff morale and are used in lecturer employment decisions, the current study focuses on this aspect.

**Student perspectives of Student Evaluation of Teaching surveys**

Student Evaluation of Teaching surveys, from the students’ point of view, are voluntary. The questionnaire includes a prescribed number of core items (requiring responses that usually fit on a five-point Likert scale) designed to gather student feedback on various aspects of the teaching of the lecturer being surveyed. The questionnaire also includes a section in which students may make additional comments on issues not necessarily covered by the Likert scale questions. The data gathered may be used to support a lecturer’s probation, performance review or application for promotion. Hence, the accuracy and reliability of this data is paramount. It will be accurate and reliable only to the extent that respondents clearly understand the true intent of the questions, and if they choose to answer them freely and without bias. But whether students meet either of these conditions is questionable, which casts accuracy and validity of data into doubt. This issue is closely examined in this section.

The voluntary nature of Student Evaluation of Teaching surveys is problematic. We argue that the voluntary survey process has a strong likelihood of generating a ‘voluntary response’ bias that results from the over-representation of individuals who have strong views in the participating group (sample). Conversely, students who are generally satisfied with the lecturer may choose not to participate in the surveys, believing there is little point in adding yet another response about normal, acceptable course content and delivery. This drop-out behaviour is in the nature of what economists call the ‘free rider’ problem. The relative hollowing out of the middle ground leaves extreme students in the sample. But the extremes are not equal. The urge to participate is likely to be greater for students who are not satisfied at all than those who are highly satisfied. This means there is a greater than zero probability that, in overall terms, the data gathered is not a true reflection of a lecturer’s behaviour or teaching method as they might have been distorted by individuals with strong negative opinions unrelated to the weaknesses or deficiencies in lecturer ability or teaching method.

Another issue of concern is that a majority of students may not have matured enough to have a sense of obligation to do the right thing and take ownership of the actions they have personally taken (or not) with respect to study. It is possible they may not have a reasonable idea of how (if at all) their actions and decisions impact on others, or at the least, lack any appropriate knowledge or understanding of the Student Evaluation of Teaching surveys process and its importance. Worrell and Kuterbach (2001) agree with this assumption in that they argued against using the student survey as a stand-alone measurement tool because students do not have adequate knowledge of sufficient criteria for observing or understanding lecturer effectiveness.

The nature of the relationship between the lecturer and students, which can be characterised as being based on an implicit contract, is another element that can influence students’ survey intentions and motivations. On the assumption that individuals are rational, each party in a contract is expected to make an effort to benefit most by attempting to maximise rights and minimise obligations. In the implicit contract, such rights and obligations are not spelt-out stipulations but are perceived only; and thus are open to individuals’ subjective judgment. With respect to the lecturer (who is likely to be aware of fiduciary obligations) and students, there can exist a significant gap between student expectations and the lecturer’s delivery. This expectations gap can apply to all sorts of engagements including assessment of student work, and provision of feedback. In the absence of any clear stipulation, students may perceive the fiduciary obligation to be much more liberal and broader in scope than what lecturers think. The expectations gap can cause friction resulting in students unfairly rating lecturers.
Management perceptions of Student Evaluation of Teaching surveys

How university management views the data collected through the Student Evaluation of Teaching survey process is also worth considering, given the likelihood of bias. Psychologically, humans in general tend to display a significantly greater reaction to negative information than to positive information. This is likely to apply in the context of Student Evaluation of Teaching survey data. For example, one respondent's critically negative assessment can significantly affect the management's overall perception of the lecturer concerned, even if all other respondents have been generally positive. This principle also applies to the average response to an individual question in a set of questions with generally positive average responses. That is, the one negative average response overshadows all the others. Arguably, this is a perversion of the concept of management by exception; the concept widely used in the management accounting.

Management preconceptions also play a role. Generally, a positive response is expected and, as such, management may not even recognise it. On the other hand, a negative response, even if it is just one alone, attracts the management's attention instantly. While this perception may vary from individual to individual, this is likely to exist in all individuals and its potential influence on how Student Evaluation of Teaching survey information is used certainly raises some concerns.

Most important of all, the data gathered through Student Evaluation of Teaching surveys is not verified or followed up. That means, management takes them at their face value. Given the categorical (ratings) nature of the survey questions, respondents might not have been able to make clear what they meant in giving a particular response. Further, students’ responses may have been overly influenced by some transitory situation that impacted on their judgement at that particular time. Verification or a follow-up process would have given students the opportunity to clarify what they meant. However, this is not part of such surveys process. Accordingly, this can result in management perceiving students’ assessment of particular lecturers differently from what students had intended. This situation, again, raises the issue of the reliability of the data, and the potential for undesirable consequences should such data be used for making a decision. The critical discussion in the preceding sections is summarised in Table 1 below.

Table 1: Factors potentially affecting accuracy and reliability of Student Evaluation of Teaching data

<table>
<thead>
<tr>
<th>Student perspective</th>
<th>Management perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voluntary nature of survey</td>
<td>Human tendency to over-react to negative information</td>
</tr>
<tr>
<td>Student's level of maturity</td>
<td>Management expectations</td>
</tr>
<tr>
<td>Students’ level of knowledge</td>
<td>Lack of validation and follow-up process on data collected through the survey process</td>
</tr>
<tr>
<td>Student expectations</td>
<td></td>
</tr>
<tr>
<td>Fiduciary nature of lecturer’s relation with students</td>
<td></td>
</tr>
</tbody>
</table>

Data and research approach

Data for this study is collected from the Student Evaluation of Teaching survey responses of 74 students across four different classes at a university campus in Australia which caters education solely to international students. While the focus of the study is on Student Evaluation of Teaching, the Student Evaluation of Units survey was also conducted to explore the possibility of whether a student’s perception of a lecturer has an impact on the student’s assessment of units or vice versa. This is important since it seems to be common practice for universities to run both surveys concurrently. Table 2 below examines the sample in terms of the nationality of the respondents.
Table 2:
Nationality representation of 74 sample participating students

<table>
<thead>
<tr>
<th>Nationality</th>
<th>Counts</th>
<th>Per cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indian</td>
<td>20</td>
<td>27</td>
</tr>
<tr>
<td>Nepalese</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>Chinese</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>Filipinos</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Brazilian</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Pakistani</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Others</td>
<td>17</td>
<td>23</td>
</tr>
</tbody>
</table>

Twenty-seven per cent of students taking part in the study were from India. The next highest grouping was from Nepal with 15 per cent, followed by China and the Philippines at 14 and 11 per cent respectively. Just under a quarter of the students were from countries other than those specified in the surveys.

The four sets of surveys were conducted in normal class-time context without students realising they were participating in a study. The lecturers’ names were replaced by dummy codes on the survey forms but the lecturers’ identities were clearly known to the students since the surveys were conducted towards the end of semester. On the Student Evaluation of units forms, the unit title was similarly replaced by a dummy code. The Student Evaluation of Teaching survey asked 9 questions whereas the Student Evaluation of Units questionnaires included 9 questions. Each question had five levels of response — ‘Strongly Disagree’, ‘Disagree’, ‘Neutral’, ‘Agree’, and ‘Strongly Agree’. Students were also given the option to choose ‘Not Applicable’ where the question asked was not relevant to them. In addition, both surveys included optional comment sections for additional comments, particularly about issues not addressed by the survey questions. All appropriate standard survey procedures were followed during the survey process.

This study uses a mixed approach to analyse data and report findings. The method of analysis will mainly be a descriptive depiction of trends in variables of interest but some econometric analysis is also employed where appropriate.

Research findings

In this section, results of the study are reported. Table 3 presents results relating to our argument that students may not provide a considered response. In this table 60 (80 percent) of the 74 students fall into three systematic categories.

Table 3:
Student Evaluation of Teaching survey responding patters of 74 sample respondents

<table>
<thead>
<tr>
<th>Responding pattern</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
<th>Per cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>One type response only (i.e. ticked all boxes in a single column)</td>
<td>22</td>
<td>16</td>
<td>38</td>
<td>51</td>
</tr>
<tr>
<td>One type response with just one different response (i.e. ticked all boxes in a column except one)</td>
<td>7</td>
<td>5</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>One type response with two different response (i.e. ticked all boxes in a column except two)</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>33</td>
<td>27</td>
<td>60</td>
<td>80</td>
</tr>
</tbody>
</table>

Note: Fifty-three per cent of sample respondents were male whilst 47 per cent were female.

Fifty-one per cent of students ticked all the boxes in a single column. In addition, 16 per cent ticked all boxes in a column with the exception of one response and 13 per cent ticked all
boxes in a column with the exception of two. This pattern does not seem to be noticeably different between male and female participants. However, it is what is absent from Table 3 that is of greatest interest: that only 20 per cent of students submitted a range of responses indicative (one would hope) of a considered appraisal of each separate question.

Table 4:
The question-by-question responding patterns of sample of 74 respondents (Student Evaluation of Teaching)

<table>
<thead>
<tr>
<th>Survey questions</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>0%</td>
<td>1%</td>
<td>9%</td>
<td>41%</td>
<td>47%</td>
</tr>
<tr>
<td>Question 2</td>
<td>0%</td>
<td>1%</td>
<td>9%</td>
<td>31%</td>
<td>58%</td>
</tr>
<tr>
<td>Question 3</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
<td>45%</td>
<td>53%</td>
</tr>
<tr>
<td>Question 4</td>
<td>0%</td>
<td>3%</td>
<td>8%</td>
<td>38%</td>
<td>51%</td>
</tr>
<tr>
<td>Question 5</td>
<td>0%</td>
<td>3%</td>
<td>11%</td>
<td>35%</td>
<td>51%</td>
</tr>
<tr>
<td>Question 6</td>
<td>0%</td>
<td>3%</td>
<td>15%</td>
<td>34%</td>
<td>50%</td>
</tr>
<tr>
<td>Question 7</td>
<td>0%</td>
<td>3%</td>
<td>5%</td>
<td>34%</td>
<td>58%</td>
</tr>
<tr>
<td>Question 8</td>
<td>0%</td>
<td>1%</td>
<td>5%</td>
<td>45%</td>
<td>49%</td>
</tr>
<tr>
<td>Question 9</td>
<td>0%</td>
<td>1%</td>
<td>9%</td>
<td>32%</td>
<td>57%</td>
</tr>
<tr>
<td>Average responses</td>
<td>0%</td>
<td>2%</td>
<td>8%</td>
<td>37%</td>
<td>53%</td>
</tr>
</tbody>
</table>

As Table 4 shows, the students mostly rated their lecturers positively across all performance indicators. On average, ‘Strongly Agree’ and ‘Agree’ accounted for 53 per cent and 37 per cent respectively. The percentages of the ‘Strongly Agree’ response are higher than the ‘Agree’ response across all performance indicators. This phenomenon is explained by the results observed in Table 3 that students generally tended to choose all boxes in one column. On average, the ‘Neutral’ response received 8 per cent whereas just 2 per cent of responses showed disapproval of the lectures’ performance across the performance indicators. We did not find this pattern to be country-specific.

We turn now to the question as to whether students’ assessment of a lecturer (Student Evaluation of Teaching) influences their assessment of the unit (Student Evaluation of Units) he or she delivers particularly given both surveys are carried out concurrently. Logically, an extreme correlation between two surveys may be argued to be indicative of unconsidered subjectivity than considered analyses of the questions asked. Average responses of two surveys are presented in the scatter plot graph in Figure 1.

Figure 1:
Scatter plot graph of average Student Evaluation of Teaching and Student Evaluation of Units survey responses
Note: Each survey question incorporated a five-point Likert scale — 'Strongly Disagree', 'Disagree', 'Neutral', 'Agree', and 'Strongly Agree'. The figure shows the relation of averaged Likert scale responses for Student Evaluation of Teaching and Student Evaluation of Units per student over the sample of 74 students. (A student who provided a low Student Evaluation of Teaching rating was also likely to provide a low Student Evaluation of Units rating)

In Figure 1, the $R^2$ value is 0.9493 while the correlation coefficient between the variables (not reported) is 0.97 and significant at the 1 per cent level suggesting that the two variables are positively and significantly correlated. While it is impossible to determine causation from simple measures of association, the set of possibilities is actually relatively small. Effectively, one of three possible phenomena are driving the very close association between the two sets of responses. The first is that changes in the Student Evaluation of Units are mostly explained by changes in the Student Evaluation of Teaching, which implies a lack of thought put into the Student Evaluation of Units responses. The second is that changes in the Student Evaluation of Teaching are mostly explained by changes in Student Evaluation of Units, implying a lack of thought put into the Student Evaluation of Teaching responses. The third possibility is that the responses to both questionnaire surveys are largely driven by some outside factor. The most likely outside factor candidate is of the sort we argue earlier in the paper to be the case — students ticking boxes without putting thought into what the questions are actually asking.

The Student Evaluation of Teaching survey questionnaire also provided students an option to furnish additional comments about the lecturer. Seventeen respondents (approximately 23 per cent) provided comments. However, the issue here is not the low level of respondents but that they do not add any further information that may clarify their assessment of the lecturer and the unit. The comments tend to be generic, and in some cases, a single word. For example, one student commented 'excellent' while another wrote 'all good'. 'She (He) is a good lecturer' is another commonly made comment. While very few, there were some comments about the lecturer's teaching method or style of delivery. For example, one respondent commented that the 'lecturer is going fast, some topics are not clear and she should explain it more precisely'.

**Discussion and policy implications**

Student feedback plays a crucial role in identifying issues and deficiencies in material delivered and the teaching effectiveness of the person delivering it. The information should provide a valuable input for management to devise a range of initiatives such as staff professional development programmes and course review processes. When it is used in assessment processes such as probation performance reviews or applications for promotion, accuracy and reliability becomes even more vital. The results reported in this paper raise the question of how reliable Student Evaluation of Teaching survey data actually is. They suggest there is a lack of diligence on the part of students in answering the survey questions. Further, the results revealed a statistically significant direct relationship between student's responses for Student Evaluation of Teaching and Student Evaluation of Units surveys. This is another indication of systematic-pattern thoughtless behaviour.

So, what can be done? At the very least, the Student Evaluation of Teaching and Student Evaluation of Units surveys should be conducted in class on separate occasions. Second, the students should be discouraged from viewing such surveys as a time trial of peripheral relevance to be completed as fast as boxes can be ticked. Some discussion of what Student Evaluation of Teaching surveys is all about needs to be given. The timing in the course calendar is also critical. Perhaps it should be administered before students receive their coursework marks. It is also important that some of the responses, particularly that are extreme in nature, are closely examined. This may however, be difficult due to anonymous nature of survey. There is little doubt on the role student feedback plays to identify deficiencies and areas needing improvement in course structure and delivery methods; but this can be useful only to the extent that the survey process applied is a robust and the data collected is accurate and reliable.
Conclusion and potential future research avenues

This study explored the diligence exercised by students apropos of the ritual of evaluating lecturer effectiveness along with judging the value of the course unit. It argued that data collected through the Student Evaluation of Teaching surveys process may suffer from lack of accuracy and reliability and provided empirical evidence supporting this argument. We argue that this warrants further investigation.

The study, like any others, suffers from some limitations and it is important that it is acknowledged. The main limitation is that the survey observed was conducted on a relatively small sample. This resulted in data that was not overly rich in content and quality. Nevertheless, data collection involved the use of a basic survey with a set number of ranked questions.

This study opens up several future research possibilities. One may be a study on the context that also includes local students (Australian Nationals) to see whether the findings observed in an exclusively international student context is supported. An investigation across institutions with bigger samples is also a promising possibility. Another possible avenue is to use an interview method to collect data to get an in-depth insight into students' perceptions of lecturers. Delving into scenarios where student feedback for the lecturer was used in a high-stake environment is certainly another promising future research avenue.

Acknowledgements

We extend our gratitude to Dr Sidney Lung and Sevasti Tsipidis for their constructive comments. Thank you also to the students who willingly participated in the study.

References


