Burnt at the Student Evaluation Stake – the penalty for failing students

Elisa Backer
The Business School
University of Ballarat
Victoria, Australia
Phone: +61 3 5327 9645
Email: e.backer@ballarat.edu.au

ABSTRACT

Student Evaluations of Teaching (SET) are a common device for assessing units and teaching ability in higher education. Despite the wealth of research in the area of SETs, little has been done to examine student and academic perceptions of SETs. This research examined student (n=235) and academic (n=49) perceptions concerning SETs at one Australian regional university. Almost one-third of respondents felt that some students punish academics for failing their work by giving the lecturer low scores on the SET form. Thus, academics can essentially be burnt at the student evaluation stake as punishment for failing students.

Keywords: SETs, student evaluations of teaching, Dr Fox
Introduction

Student Evaluations of Teaching (SET) is a formal process in higher education that typically occurs at the end of the unit. Students are asked to fill in a survey (either paper-based or online) to provide feedback on the teaching staff and unit. Numbers are allocated against each question to provide a ‘rating’ of how good the unit was, and how good the lecturer/tutor was. There is an abundance of literature that discusses the value of this process, and equally in abundance exists literature highlighting the deficiencies of the system.

Despite the extensive research in the field, student and academic perceptions have received little attention. In terms of student perceptions of SETs, “implicit in all of the literature is the assumption that students answer these anonymous instruments honestly and willingly” (El Hassan, 2009, p.320). As such, how honestly the students answer the questions has been virtually neglected. Similarly, academics’ perceptions on SETs have not been well researched (El Hassan, 2009). In particular, examination of student and academic perceptions of SETs has been overlooked in Australia.

This research aimed to contribute to the limited body of literature that has examined the area of student and academic perceptions of SETs. Specifically, this research aimed to consider whether students and academics at this one university believed that SETs were reliable, and whether students used the process of evaluating their lecturer as an opportunity to punish them for giving them low grades.

Literature Review

Student Evaluations of Teaching (SET) are a common device for assessing units and teaching ability in higher education (Pounder, 2007). In 1915, the first paper discussing the teaching rating scale was published (Denson, Loveday, & Dalton, 2010). The benefits and drawbacks of SETs have been discussed as far back as the 1920s (Wachtel, 1998), with mixed reports on the validity of the instruments and their usefulness in determining instructional effectiveness. Some writers have expressed reservations over the validity of SETs (Chandler, 1978; Dowell, 1982; Powell, 1978; Sheehan, 1975; Slade & McConville, 2006; Small, Hollenbeck & Haley, 1982; Vasta & Sarmiento, 1979; Wilson, 1998). In particular, this form of evaluation has been criticised as encouraging inflating grades and “dumb(ing) down material” (Wilson, 1998, p.A12) in order to receive higher scores. Particular criticism on the emphasis that universities place on SETs has also been provided, drawing upon the numerous flaws in the instrument (Gray & Bergmann, 2003). According to Gray and Bergmann (2003) 72% of academics stated that the over-reliance on SETs tends to encourage a watering-down of course content. They also claim that significantly better ratings are achieved by better looking teachers (Gray & Bergmann, 2003). Race and gender can also affect student evaluations (Merritt, 2008).

Gray and Bergmann (2003) claim that universities tend to assume that the average of the student evaluations suggest that a teacher is bad at their job if their scores fall below that average. They point out that “if all those who were below average were fired, the average would simply rise, and about half the previously “good” teachers would then be below the new average, miraculously reborn as “bad” teachers” (Gray & Bergmann, 2003).

However, SETs have also been discussed as being valid and accurate (Centra, 1977; Cohen, 1981; Koon & Murray, 1995; Marsh, 1984; McKeachie, 1990; Murray, Rushton & Paunonen, 1990; Ramsden, 1991; Seldin, 1993). Regardless of whether SETs are valid and accurate, or flawed, they do offer benefits. Student evaluation feedback can
be an important product in striving for continual improvement (Wachtel, 1998). Used
in the right way, student feedback can be valuable for an academic who can then
reflect on how they might change elements of a unit or its delivery to improve
teaching and learning.

This argument is used to justify the continued use of the mandatory student feedback
questionnaires by many of the higher education institutions. Purportedly this is done
to allow those institutions to gauge feedback on unit and teaching performance at or
towards the end of each semester. They are widely used in Australia, the UK, and the
US (Denson, et al, 2010). SETs are only one of the survey instruments that students
may come across in the duration of their studies. They may also be asked to fill in
Student Evaluations of Unit or Course (SEU / SEC). In Australia they may be asked to
complete a Course Experience Questionnaire (CEQ), which is a national questionnaire
designed to compare graduate experiences between institutions in Australia.

Whilst recognising the importance of the CEQ, it is not designed to provide feedback
on individual teachers or individual units. Hence, the role of a SET is clear. Although,
the validity of any of these surveys can be questioned, and whether there is any link
between results and improved teaching or curricula can also be queried.

One criticism of SETs is that it is conducted at the end of the semester, and as such
students will be unlikely to know whether the feedback they have provided will be
actioned for the next cohort of students (Tucker, Jones, & Straker, 2008). This lack of
feedback may result in creating a “climate in which students do not take the existing
feedback mechanisms seriously but, rather, may use the opportunity to express their
frustration, in general, rather than providing constructive feedback” (Tucker, et al,

Interestingly, recognising the limitations of the evaluation process, Curtin University of
Technology’s School of Physiotherapy developed an evaluation system for units, which
was aimed at improving the CEQ (Tucker, et al, 2008). This included ensuring the
feedback loop for students was not closed. The feedback was made public. Normally
only the academic responsible for the unit receives the feedback and there is no
support and follow up afterwards. However, at the Curtin University of Technology’s
School of Physiotherapy, all academics and students knew of all scores, which created
an open and transparent system and a sharing of teaching concepts (Tucker, et al,
2008). It was this that Tucker et al (2008) attribute their CEQ success to. Their CEQ

Therefore, it is possible that the process of evaluations is incomplete unless the
institution does more with results. However, the survey instrument itself can still be
limited as many of the questions tend to focus on the teacher rather than the script
(Biggs, 2003). Therefore, SETs tend to measure charisma (the charm and magnetism
of the lecturer) as opposed to content (the coverage of critical issues by the lecturer)
(Ware & Williams, 1975). SETs have also been blamed for encouraging some
academics to lower their standards and raise their grades in order to “teach to the
evaluations” (Wilson, 1998, p. A12). With SET scores being used in tenure and
promotion decisions, and academics believing that course difficulty will bias those
scores, the temptation to reduce difficulty or raise grades exists (Centra, 2003).

However, whilst Centra (2003) found that difficult courses scored poorly on SETs, so
too did courses that were considered too easy. In fact, the teacher’s personality has
been found to be a key issue in affecting course grades (Marsh, 2007). Research by
Marsh (2007) showed that teaching does not improve over time and that “SETs are
primarily a function of the teacher who teaches a course rather than the course that is
taught” (p.776). The value of the academic’s enthusiasm in influencing students was
highlighted in research that sought to examine whether students are motivated by
mandatory attendance policies (Verbeeten, 2007). When asked “what can your professor do to make you come to class?” (Verbeeten, 2007, p.31), the most popular response (90.4%) was “if he/she makes the class interesting” (Verbeeten, 2007, p.31). The second most common response (60%) was “if he/she is excited about the topic” (Verbeeten, 2007, p.31). This suggests that there is value in adding charisma to engage students in the learning environment. However, as SETs do tend to focus on measuring charisma, SETs should not be relied on exclusively.

Critics of SETs have seized on Naftulin, Ware, and Donnelly’s (1973) research, which is commonly referred to as the Dr Fox study. In the study, the authors hypothesised that SETs largely rate charisma and popularity and that even experienced educators exposed to an irrelevant and meaningless but charismatic lecture can be seduced into feeling satisfied that they have learned. To examine this, the authors prepared a professional actor “who looked distinguished and sounded authoritative” (Naftulin, Ware, & Donnelly, 1973, p.631) to deliver a charismatic but insubstantial lecture on a topic that the actor had no knowledge of. Introduced as Dr. Myron L. Fox, as being an authority on the application of mathematics to human behaviour, he delivered a one-hour lecture on ‘Mathematical Game Theory as Applied to Physician Education’. As per his coaching, he deliberately used an excessive amount of “double-talk, neologisms, non sequiturs, and contradictory statements…interspersed with parenthetical humor and meaningless references to unrelated topics” (Naftulin et al, 1973, p.631-632). In what has become known as “the Dr Fox effect”, the respondents responded favourably to the lecture through feedback provided by an anonymous questionnaire. Across the three groups of respondents, 91% of the 55 respondents believed that Dr Fox stimulated their thinking. Despite Dr Fox having no knowledge of the subject he lectured in, the respondents were seduced into feeling that they had learned.

Whilst the Dr Fox study (Naftulin, et al, 1973) has been criticised in terms of methodology (Marsh & Ware, 1982; Wachtel, 1998) “the most serious charge levelled against the Dr. Fox study is irrelevance” (Kulik, 2001, p.18). Kulik (2001) points out that the study does not align with the manner in which higher education teaching and ratings are done, since Dr Fox only presented one lecture before being rated. The point that Kulik (2001) makes on this issue is that whilst Dr Fox may have been able to make a favourable impression to his audience after one lecture, “surely everyone would have caught on to the fraud if Dr Fox were the lecturer in a semester-long course” (Kulik, 2001, p.18).

With SET scores playing an increasingly critical role in recruiting and promoting academics, the question about their value and how an academic should use the feedback provided remains. Particularly if no dialogue between the academic and the institution occurs in response to the scores, as it does at Curtin University of Technology. Research into SETs is substantial, outlining the important role that the lecturer’s personality has in teaching any course. It also highlights the perception that course difficulty will bias the results. However, given that students are aware of these scores being important, what is not known, is whether students use their feedback as an opportunity to punish a lecturer for low grades, particularly if there is no feedback system as suggested by Tucker et al (2008).

Whilst the area of SETs has been extensively explored for almost a century now, gaps are still evident, and the subject area has attracted little interest within Australia. One such gap is the perception of the SET survey instrument by academics and students. Literature has already highlighted that SETs should not be used in isolation and that it is only one type of evidence. The problems associated with using it at all has been relatively ignored, with the exception of the 1973 seminal study into the “Dr Fox” effect “ (Naftulin et al, 1973). SETs remain a popular tool for use in promotion and recruitment applications in higher education. However, the potential power that is offered to students in providing formal feedback is an important issue to consider. In
knowing how much effort the student puts into providing feedback, and whether students use the feedback system as an opportunity to punish, feedback can be put into perspective and allow for more accurate reflection by the practitioner. Therefore, this research aims to contribute to theory and practice by specifically examining how academics and students in one tertiary institution within Australia perceive SETs.

Method

In order to gauge academic and student perceptions of the SET survey instruments, students and academics in one institution were surveyed separately. Students were surveyed during late 2009. The survey of undergraduate students occurred at one Australian regional university where paper-based SETs are conducted. Lecturers of undergraduate units who were planning to undertake SETs / SECs for their unit were asked if they would allow a short survey to be undertaken after the SET survey had been conducted. The rationale behind this was that if students had only just filled in the SET they would have the perceptions of this fresh in their minds. As the students would be asked to fill in an additional survey, the questionnaire was deliberately designed to be very short and simple and hone in on vital questions only. Students who had previously filled in the survey were ineligible to respond. Students across various disciplines were captured, including accounting, economics, management, tourism, marketing, law, entrepreneurship, and commerce as well as smaller numbers of arts students and human movement students. The number of responses was 235, and these results were entered into SPSS. A notation in SPSS was made regarding the class (discipline and year level) the responses were collected from.

Surveying of academics occurred in early 2010. A short online survey was developed through 'Survey Monkey'. An announcement inviting academics to participate in the survey with a link to the online survey was posted on the academic communication section of the university website. In addition, an email was sent to the heads of each school asking for their support in forwarding the email to their academic staff, with the link to the online survey on the email. The number of responses was 49.

Results

Student survey results

The results from the student survey are presented first. The first question asked whether students felt that the current SET survey instrument accurately captured students’ views of the teaching. The options for selecting were disagree, neutral, and agree. Almost half (47.5%) of the respondents were neutral on this and only around one-third of the students agreed (34.7%). There were 17.4% that disagreed. The results from the first question were then compared to the different classes where the surveying took place (ie tourism, law, economics with year levels). A chi square test revealed no significant difference (at the 95% confidence level) between the responses provided with respect to what class the surveying took place in.

The second question asked whether students felt that the task of providing feedback via the SET system was enjoyable. Around one-quarter (27.1%) disagreed. Around one-third (31.4%) agreed about it being enjoyable. Neutral responses were received by 41.5% of respondents. The results were then compared to the classes where the surveying took place. A chi square test revealed no overall significant difference (at the 95% confidence level) between the responses provided with respect to what class the surveying took place in. The only difference found was for a first-year human resource management class. A z-test revealed a significant difference between those that were neutral (52%) compared with those that disagreed (12%).
The third question asked whether students felt that the task of providing feedback via the SET system was helpful for the teacher (Table 1). More than half (56%) considered SETs as a valuable process for the teacher. Around one-third were neutral about this and less than 10 per cent disagreed, feeling that it was not helpful for the teacher. The results were then compared to the classes where the surveying took place. A chi square test revealed an overall significant difference between the responses with respect to what class the surveying took place in. Contributing to this overall difference was a significant difference between the neutral and agree responses for the first-year tourism class. A z-test revealed a significant difference between neutral responses (54.1%) and responses from those who agreed (37.8%).

**Table 1:**

*Perception of whether the SET is helpful for the teacher*

<table>
<thead>
<tr>
<th>Q. 3 The current SET forms are helpful for the teacher (n=230)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree</td>
</tr>
<tr>
<td>8.3%</td>
</tr>
</tbody>
</table>

$\chi^2 = 26.602, df=16, p<0.05$

The fourth question asked students whether some students provide low ratings to teachers who had given them low grades (Table 2). The results showed that respondents were reasonably equally split between the three response options, and only around one-third of students disagreed. Almost 30% of students agreed that some students do provide low ratings to teachers as punishments for low grades. A chi square test revealed an overall significant difference (at the 95% confidence level) between the responses with respect to what class the surveying took place in. This difference was largely driven by the responses for third-year accounting. A z-test revealed that the agree responses (71.4%) were significantly higher than the neutral (19%) and disagree (9.5%) responses. The unit with the second highest response rate to the agree option was second-year law (40%), followed by second-year accounting (36.4%), and first-year tourism (35.1%).

**Table 2:**

*Students provide low ratings to teachers as punishment for low grades*

<table>
<thead>
<tr>
<th>Q. 4 Some students provide low ratings to teachers as punishment for low grades (n=234)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree</td>
</tr>
<tr>
<td>34.3%</td>
</tr>
</tbody>
</table>

$\chi^2 = 32.865, df=16, p<0.05$

The fifth question asked students how much thought and consideration they provide when filling in the SET forms. Almost half (49.2%) of the respondents stated that they had put in “a fair amount” of thought when responding to SET forms. However, 24.6% felt that they only put in “some” effort whilst 14.4% put in “a little” effort. Only small numbers were on extreme ends of the scale with 1.7% putting in no thought and 9.3% putting in “a great deal” of thought. A chi-square test at the 95% confidence level revealed no significant overall difference when factoring in the different units and year levels of classes where surveying was undertaken.

The sixth question asked about the numbers of lectures the students had attended. This was designed to gauge how qualified students are to provide feedback on teachers and courses. Clearly, a student who has scarcely been to class is not in a position to comment on the quality of the teaching and the content of that subject, yet
they are given equal opportunity to provide their comments. The vast majority of respondents (92.3%) claimed to have attended most of the lectures, missing up to four lectures at most. This is not consistent with anecdotal information from lecturers. The number of students who were present in the classes were compared to enrolment data for those classes. The response rate was a small portion of the enrolment data, particularly for some of the disciplines. This suggests that the survey attracted a natural bias towards those who regularly attend most or all classes. A chi-square test at the 95% level revealed no significant difference when assessing different units and year levels for the responses provided.

The final question was an open ended question asking students to provide further comments if they wanted to. Whilst few respondents took this opportunity, some comments were made that reinforced the deficiencies associated with relying heavily on SET scores:

“It’s difficult to assess a lecturer as the end of the semester. It’s hard to remember everything that has happened.”

“I would think students don’t really think about their responses when filling out a SET. They just tick the boxes and don’t worry about the feedback, therefore it probably doesn’t provide that teacher with extensive feedback.”

“The thing with the SET’s is the question where it asks if we believe the teacher is interested in improving their teaching. I believe that question isn’t warranted as it’s too hard to answer. It comes down to your opinion of the person not their teaching.”

“Results or information on improvements made to units would be nice to see. It would be good to see our feedback is used.”

**Academic Survey results**

The results from the academic survey are presented next. The first question asked academics ‘Do you undertake SETs & SECs for your subjects?’. Since at present, no online system for undertaking evaluations exists, it is up to the individual academic to order and arrange the conduct of SETs / SECs. Whilst policies are in place in The Business School that state SETs / SECs must be conducted for certain units, not all academics adhere to these rules. The response sample size for this question was 48, with 50% of the respondents stating that they conduct SETs and SECs ‘for all subjects all the time’. The second most popular response was ‘only some subjects – it depends on time’ with 22.9% selecting this option. Some respondents (14.6%) selected ‘other’ and two main themes arose from this. One theme that arose was that respondents would undertake SETs / SECs every two to three deliveries. The other theme that arose was that some respondents stated that they refused to undertake SETs / SECs and had not engaged in the process for years. Reasons for this were based on validity. They felt that results were flawed and meaningless so would not partake in it. There were 8.3% of respondents that stated that they conducted SETs and SECs but ‘only the subjects I have to’. The least popular response was ‘only the subjects I have to, as well as the subjects that yield higher scores’ (4.2%).

Academics were then asked whether there had been times when they had decided not to undertake SETs and SECs because they had felt that the feedback for a unit would be lower than usual. Whilst the clear majority (79.6%) stated that they had not done this, just over one-fifth (20.4%) admitted that avoiding undertaking SETs / SECs due to anticipated low feedback. This highlights an important point about the fear of the scores reducing willingness to obtain feedback. Whilst academics may be perfectly willing to receive feedback and reflect on this, when there are scores attached, it can
breed reluctance. This can be particularly the case when the scores can be central to determining teaching ability.

The third question was designed to seek academics’ opinion on whether SETs / SECs should be compulsory for every unit over each delivery. The majority (75.5%) of the respondents felt that evaluations should be compulsory for all units for each delivery.

The next question asked of academics was based on the format of the SETs / SECs, which are presently paper-based. Questions four asked ‘do you think it would be better to have an online SETs / SECs process rather than paper-based? Respondents were roughly equally divided in their opinions with 52.1% stating yes and 47.9% saying no.

The fifth question was designed to assess how academics perceive the SET / SEC survey instrument itself (Table 3). There were five options to select from ranging from very poor up to very good. The most common response was ‘neutral’ and the average score was 2.52. Just over one-quarter (27.1%) of respondents rated the SET / SEC survey instrument as ‘very poor’. In total, 50% of respondents (n=24) rated the instrument as either ‘very poor’ or ‘poor’. Only 8.3% of respondents provided a rating of ‘very good’.

Table 3:

<table>
<thead>
<tr>
<th>Q. 5 How would you rate the value of the SET / SEC instrument to accurately evaluate the lecturer’s skills as a teacher? (n=48)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very poor</td>
</tr>
<tr>
<td>27.1%</td>
</tr>
</tbody>
</table>

The final question in the survey was simply an open-ended question. It asked ‘What is your opinion about undertaking SETs and SECs? Please expand upon your responses to the previous questions.’ This was the only question that had been designed as compulsory, thereby forcing respondents to provide some comment before ending the survey. Interestingly, most people left very detailed responses. The most common length of the response for this question was around 130 words. Only one person left a one-worded comment. Two people left particularly detailed comments (one was 453 words; the other was 443 words).

A strong theme running through the comments involved the SET survey instrument being “flawed”. For example, “The instrument is flawed, though I recognize that finding one that is not flawed may be problematical.” Similarly, “…it is well known how easily the results can be manipulated. It is risible to imagine that they have any statistical validity and unprofessional to use them as measures of staff performance.” Also discussing the flawed process was this comment:-

Evaluations of teaching are important, but the current SETs tend to be a rating of popularity…I know plenty of staff who “dumb down” their content and assessment tasks to score higher on SETs. This misses the point, but given the heavy reliance on such scores for promotion, it is understandable.

Other respondents spoke of ‘student fatigue’ associated with undertaking so many evaluations and losing too much class time because the surveys are conducted during class. There was also a theme of fear of what is done with the information by the institution. For example:-

© e-JBEST Vol.6, Iss.1 (2012) 8
I admit to destroying some unfavourable SET data at a previous university for a variety of reasons, including that my sessional employment would have been jeopardised by the negative feedback and that the subject I was handed to teach was dreadfully designed and guaranteed to get low SET/SEC results - not my fault.

A similar comment was “I don’t undertake SETs/SECs any more. There is no penalty for not doing them, but there is for not reaching some arbitrary number, which becomes permanently etched on our employment records.” These comments demonstrate a fear of what happens if the SET / SEC score is not high and the avoidance strategies undertaken to escape the perceived penalty.

**Discussion**

This research has added to the small body of literature that has examined the student and academic perceptions of SETs. Certainly anecdotal criticism from academics on SETs is well-known, however little empirical research has been undertaken to examine this aspect (Schmelkin-Pedhazur, Spencer, & Gellman, 1997). The focus in the extant literature has been placed on the fault of the instrument itself and that the emphasis is on charisma more than content. The validity and reliability of the ratings received from SETs has remained a popular focus in the extant literature (Schmelkin-Pedhazur, et al, 1997).

The focus on the problems with SET reliability has tended to result in criticism surrounding the tendency for some lecturers to teach to the evaluations and make their units easier to pass (Centra, 2003; Marsh, 2007; Wilson, 1998). However, while academics may complain about the SET process, this does not change the high importance placed on the students’ ratings of the teachers when it comes to promotion, tenure, or external job applications. In Australia and indeed other countries such as Canada, the United Kingdom, New Zealand and the USA, many universities will request proof of good teaching via SET scores.

With only around one-third of students feeling that students do not punish teachers for low grades, it means that a majority either feel that this happens or that it possibly might. This raises further weight to the existing research in regards to the validity of SET scores. Previous research has largely focused on the bias due to the personality of the individual instructor, class size, and subject type. Fewer studies have considered the punishment component that SETs may be associated with. One study that discussed this aspect reported that 40% of students admitted they knew of students scheming revenge on a lecturer through collectively providing low SET scores (Jacobs, 1987 cited in El Hassan, 2009).

Fear of such punishment by students has been associated with academics ‘dumbing down’ material to try to improve their scores, which may be needed for their next promotion. The problem has mainly arisen due to SETs being used in isolation rather than as one amongst other ways of assessing teacher ability. In a sense, “students become the primary determinant of a lecturer’s success or failure in his or her academic career” (Emery, Kramer, & Tian, 2003). This point was also raised by Comm and Mathaisel (1998) who drew an analogy with industry. As they pointed out, there are times when in certain industries, workers may evaluate their bosses, however, this will only be one of several methods that are employed to determine the ability of a manager; not the sole criteria (Comm & Mathaisel, 1998). Yet, in universities, SETs are typically the only measure of a teacher’s ability.

With such power bestowed to students, it is easy to understand how they may ‘gang up’ on a teacher and give them low SET scores as punishment. They perhaps understand that “an entire career can be terminated by not-good-enough evaluations”
(Gray & Bergmann, 2003). Essentially, academics can be burnt at the student evaluation stake, as punishment for failing students. However, if academics lower standards to protect themselves, this raises a problem for industry, in that the students may not be well prepared as future employees (Comm & Mathaisel, 1998). In such a situation it will be “costly for businesses if they must re-educate ill-prepared new employees” (Comm & Mathaisel, 1998, p.164). Another problem can arise due to failing university entry levels. If a proportion of a cohort is below the required skills, this may result in a higher failure rate. If a lecturer is likely to be questioned by management for having high fail grades, there may be an temptation to raise grades to avoid the questioning.

This perceived concern was detected through the detailed comments left in the last question on the survey of academics. Academics reported avoiding engaging in the SET process to avoid having a ‘score’ attached to their record. In one reported case, an academic destroyed the evidence.

**Conclusion**

These findings have been valuable for several reasons. Firstly, this research has contributed to the limited literature that has examined student and academic perceptions of SETs. More so, with much of the extant literature derived from the United States, this study presents a view from an Australian institution.

This research is limited is several ways. It is based on one university, and therefore may not reflect broader perceptions. Also, the samples sizes may be considered small. However, given that evaluations are not compulsory at the university in which surveying took place, the response rates are regarded as representative for the institution.

Whilst these results are applicable to one institution in Australia, the findings are considered relevant on a broader level and will have interest to any tertiary institution that operates SETs. This study has shown that students, not just academics, do not have a high opinion of the SET system. It has shown that many students do not think that it accurately captures their thoughts on the teaching and that many students do not enjoy filling in SETs. They typically think that the beneficiaries are the teachers themselves. However, based on the perceptions of academics at the same institution, many academics do not feel that they are beneficiaries; but see themselves as victims.

In this study, 30% of students stated that they felt teachers are punished for giving low grades by way of receiving low student evaluations scores. This was similar to the findings by Jacobs (as cited in El Hassan, 2009) that 40% of students reported to plotting revenge via the SET process. These findings provide some validation to the concerns raised by Tucker et al (2008) that students become frustrated with the process of SETs and do not use it for the intended manner. Punishment by way of receiving low scores could potentially have a significant effect on a teacher’s record. Therefore, it is recommended institutions become aware of the prospect that teachers may be punished by the students by way of low SET scores. Students should not be overly empowered with this charge and therefore in order to avoid burning teachers at the SET stake, procedures need to be taken into account to avoid an over-reliance of the SET.

One possibility is for peer review to be encouraged, which provides another layer of feedback, particularly for those teachers who have received low scores. Those reviews could be used in conjunction with SETs. Another possibility is for teachers to operate mid-semester feedback sheets, which could provide feedback on the unit and the
teaching prior to assessment grading being released, which may provide another form of checking against final SET scores. A third possibility is for SETs to be run earlier in the semester. Instead of operating SETs at the end of semester, if they are conducted three-quarters of the way through, students would have had ample opportunity to judge the quality of the teaching but are unlikely to have undertaken a major assessment. This may reduce low student grades influencing teacher ratings. Another option is to adopt the Curtin model and make the system more transferable and open, which may ensure students are more likely to approach the task with the right attitude, although there is no evidence to support this would occur. A final possibility is to take away the scores. The purpose of SETs was to provide the instructor with feedback to enable reflective practice and improvements. The moment numbers and ratings are associated with this, the process changed from one of feedback to one of competition and empowerment. It may be that the simple act of preventing scoring is all that is needed without reducing the benefits of the feedback sheets.

However, despite the criticisms of the SET system, it still provides teachers with a system in which feedback is obtained. As such, the important role that SETs offer must be stressed. The key point that this research shows, is that it may not be appropriate to use it in isolation.

References


