This report is a literature review, review of the current Module Evaluation Questionnaires (MEQ) and proposal for revisions to the MEQ.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>2</td>
</tr>
<tr>
<td>Purpose of SET</td>
<td>2</td>
</tr>
<tr>
<td>Reliability and validity</td>
<td>4</td>
</tr>
<tr>
<td>Discussion</td>
<td>7</td>
</tr>
<tr>
<td>Conclusions and Recommendations</td>
<td>7</td>
</tr>
<tr>
<td>Appendix</td>
<td>9</td>
</tr>
<tr>
<td>References</td>
<td>11</td>
</tr>
</tbody>
</table>
Introduction

Student Evaluation of Teaching (SET) remains a current yet controversial topic in higher education research and practice with many participants questioning the validity and reliability of SET results (Ory, 2001). Research on SET has thus far failed to provide clear answers to several critical aspects concerning the validity of SET (Spooren et al, 2013).

SET may be used for two purposes: institutional and local. Institutional SET is what university set the standard questionnaires and distribute them to all schools and departments, which is considered more traditional. The majority of SET research appears to focus on institutional level SET. Institutional SET is also often relied upon to evidence cases for promotion and is an important element of a teaching portfolio.

Local SET, on the other hand, usually refers to what each school or department set its own questionnaires in addition to the Institutional SET. Local SET is implemented as a compliment to the institutional SET providing insight and feedback on specific disciplinary questions. This can lead to confusion as students may not understand the different purposes of institutional and local SET. In this report, we review the current Sussex SET practice and propose for revisions to the Module Evaluation Questionnaire (MEQ).

Sussex SET is conducted by means of MEQ which is an important source of student feedback on teaching and learning. It belongs to an institutional type (we call it ‘University MEQ’) as it is set by the university institution-specific questions. However, many schools and departments also run local MEQs (we call it ‘local MEQs’). A major shortcoming of the University MEQ is that the current MEQ suffers from low response rates thereby reducing its usefulness and validity. In response to the limitations, local practices have grown to address the need for student feedback at a module level, however they may not be consistent year on year or across the university.

This report consists of two main sections. The former is the literature reviews and the latter is the discussion and recommendations for consideration. Literature reviews are on the purpose, validity of MEQ. The report starts with the literature review summarising the purpose of SET.

Purpose of SET

Introduction

Student Evaluations can take many different forms. The focus of this report is the summative surveys which run at the end of a term or level of study. This section outlines the stakeholders for the SET and the various purposes.

Stakeholders

There are three groups of stakeholders that could to be considered to be affected by the SETs.

Students – engaging students in the process is critical to the reliability of the data collected though the SET process. Shortcomings can be that students are overwhelmed with completing multiple SETs simultaneously, they don’t appreciate how the information is used and feel it is a pointless exercise(Bassett, et all. 2015).
University – May often have ulterior motives in the student feedback process. They might be thinking in terms of league tables – for example with the NSS. Some may not see this as a developmental tool, but will be approaching it as a KPI. This then might feed into management decision making, such as hiring, firing, promotion, pay, rewards, etc. The problem here is that a KPI may mask the real reason for the decision, which could be something subjective. (Salmon, 2009, p: 77)

Staff – The content of student SETs often misses the ‘tacit’ knowledge contained in professional practice. The questions simply might not be relevant to helping staff develop themselves.

The purpose of SET
Various purposes are identified in the literature, but the purpose of SET appears to be summarised primarily as: teaching, academic promotion related reasons and institutional accountability.

Teaching
The purpose of teaching is used for giving feedback for individual academic staff engaged in teaching as well for supervisory purposes of teaching for the School unit/course coordinators and Head of school.

Academic promotion
SET has been used for academic promotion decisions (Becker & Watts, 1999) and professional career progression as an input for appraisal exercise for tenure/promotion decisions as a measure of quality monitoring, administrative policymaking (Penny & Coe, 2004). In fact, 62% of the academic staff feel that department heads and deans make proper use of SET reports (Beran & Rokosh, 2009). The majority of institutions have been using SET for promotion since the 1970s when evaluations came into use for faculty personnel decisions in hiring purposes (Becker & Watts, 1999). It is implied that the close link among SET, academic promotion and HR department.

Accountability
Accountability appears to be considered of paramount importance at universities not only due to the British government emphasis on university’s accountability and transparency across high education sector (2017), but also accountability links to attractive concepts such as ‘quality’, ‘autonomy’ and ‘integrity’, all of which give stakeholders special attention and positive connotation (Stensaker & Harvey, 2011). The universities report to stakeholders regularly and openly (transparency), documenting (visibility) on the outcome of their effective teaching, research activities and institutional management. Institutional accountability consists of the myriads of either tangible or intangible expectations (Kearns, 1998) from the diverse stakeholders such as students, staff, politicians, alumni, funding bodies and the general public (Kearns, 1998). “Students as consumer or partners” (Bienefeld & Almqvist, 2004) and the concept of “consumer satisfaction” (Blackmore, 2009; Olivares, 2003; Titus, 2008) has brought into education and that is how students play an important part in evaluating tutors.

SET can offer evidence of: teaching quality for individual tutors (Spooren et al., 2013) and institutional internal quality-assurance accountability. Evidence of internal quality-assurance
processes provides institutional accountability (Johnson, 2000). As current emphasis on SET data is on quantitative rating (Likert-scale questions), the easiness to collect, analyse and present the SET data (Penny, 2003) also contributes why the majority of institutions use SET as evidence.

Reliability and validity

To fulfil their potential the SET should meet tests of validity and reliability. We examine both external and internal validity against current Sussex MEQ (institutional and local).

Threats to external validity

The definition of external validity is “the extent to which samples, settings, treatment variables and measurement variables can be generalised beyond the study” (Gliner et al., 2009, 128). Population external validity is one aspect of external validity which may be threat to validity with regards to University MEQ. Population external validity (Gliner et al, 2009) relates to the sample/population. When we examine population external validity, both University and local MEQ results may not be the exact representation of the Sussex University students which include ethnicity. To address this, information on students’ ethnicity should be included in the University MEQ.

Another threat to external validity is ‘external replication’ (Graziano & Raulin, 2000). It is a concept which is often discussed in conducting research as a threat to external validity. Evaluating current MEQs (both University and local), University MEQ may not be threats to external validity, but local MEQ results may be threats to external validity, as local MEQ may not be necessary to replicate externally to other schools. However, local MEQ are useful for the tutors’ feedback and could be used within the school.

Reliability

Reliability is defined as “stability or consistency with which we measure something” (Robson, 2002, 101) and “for a research instrument to be reliable, it must be consistent” (Gray, 2004, 173). Therefore, the consistency is considered important in conducting both University and local MEQs. If the implementation of local MEQs is consistent across university, local MEQ results also become comparable among disciplines. However, we have to take consideration of bias as a source of unreliability which threatens validity. Bias occurs from tutors, students and other factors, which will be discussed next.

Tutors’ related bias

Two biases related to the tutors’ grading affect validity of SET: validity hypothesis and grading-leniency hypothesis. Validity hypothesis is defined as “students who have learned more in the class will receive higher grades and will naturally rate the professor more highly because of the knowledge they have gained in the course” (Patrick, 2011, 241). However, Abrami, Dickens, Perry & Leventhal (1980) argue that the effect of grading standards on students’ ratings is very small and inconsistent across different rating items.

Grading-leniency hypothesis is defined “the professor’s leniency in assigning grades favourably influences student evaluation scores” (Patrick, 2011, 241). Simply, tutors who give higher grades also receive better evaluation (Carrell & West, 2010; Johnson, 2003;
Weinberg, Fleisher & Hashimoto, 2009). Some researchers suggest that tutors buy good evaluations by giving high grades (Isely & Singh, 2005; Langbein, 2008; McPherson, 2006; McPherson & Todd Jewell, 2007). There is a link between SET and tutors’ promotion (Becker & Watts, 1999) and career progression i.e. input for appraisal exercises (e.g., tenure/promotion decisions). As career promotion is important for tutors, some tutors may strategically adjust their grades to please the students (Braga, Paccagnella & Pellizzari, 2014) and others may teach the test content as students might reward tutors (Braga, Paccagnella & Pellizzari, 2014).

**Students’ related bias**
The following 11 factors have been identified as contributors to student related bias:

1. Weather (Braga, Paccagnella & Pellizzari, 2014)
   Students tend to evaluate negatively on rainy and cold days.

2. Time of the day (Feldman, 1978)
   This factor has been suggested as a possible biasing factor.

3. Lower-track students
   Students evaluate their tutors based on how much they have enjoyed the course. In another words, a good tutor can get bad evaluations if they teach a class consisting of the least able ones (Braga, Paccagnella & Pellizzari, 2014).

4. Gender (male vs. female) (Bavishi, Madera & Hebl, 2010)
   Peterson (2019) suggests that relatively simple changes to the language used in SETs can make a positive impact in assessments of female professors. Yet the authors caution that if these changes were widely adopted, students (and their biases) might adjust to the new system and the positive effect for female professors might wear off.

5. Racial (black, Asian, white) stereotype by students (Bavishi & Madera, 2010)

6. Physical attractiveness (Campbell et al., 2005; Gurung & Vespia, 2007; Hamermesch & Parker, 2005; Riniolo et al., 2006)

7. Image compatibility (Dunegan & Hrivnak, 2003), initial impressions of a tutor (Tom et al., 2010)
   First impression- various studies showed that personality judgements based on only brief instances, and sometimes even without sound, proved significantly related to end-of-term evaluations of teaching (Ambady and Rosenthal 1993; Clayson and Sheffet 2006; Tom, Tong, and Hesse 2010).

8. Students’ personality traits (Patrick, 2011)
   If students have an agreeable trait, they evaluate their tutors better.

9. Tutors’ personality traits
   Tutors’ Big Five personality (Patrick, 2011) traits consist of Neuroticism (mental health),
extraversion (warmth, assertiveness), openness to experience (uncertainty avoidance), agreeableness (as opposed to ‘distrustful’, ‘antagonistic’, ‘self-centred’ and ‘compete rather than collaborate with others’), conscientiousness (goal-oriented, organised, punctual dependable).

10. Students’ prior interest- affects the students’ learning process and thus their grade expectancy. (Spooren et al. 2013)

11. Students’ feedback-seeking behaviours and anxiety

Student feedback-seeking behaviours such as asking questions in class have been shown to be negatively influenced by fear of embarrassment (Fassinger, 1995); many students are known to be hesitant or even fearful of asking questions in class (Fassinger, 1995; Ng & Ang, 1997). These anxieties may come from a general fear of embarrassment by students or, as some have found, cultural values of not wanting to lose face in front of others by asking questions that reveal their lack of knowledge (Hwang, 1987; Ng & Ang, 1997). Hence, tutors with heavy emphasis on collecting feedback from students without considering students’ fear of embarrassment, could be penalised at their MEQ by such students. Loosing face factor would be felt with Asian students. (Hwang, et.al, 2003)

Other variables which affects validity:

1. Level of the courses
   Tutors who teach more advanced courses are rated more favourably than teaching lower-level courses (Goldber & Callahan, 1991; Moritsch & Suter, 1988; Cranton & Smith, 1986; Feldman, 1978; Marsh, 1980). This is due to the nature of heterogeneity in student composition in lower level courses. Heterogeneity makes it extremely difficult for tutors to tailor to the students’ preferred teaching and learning experience to satisfy all. By attempting to suit the curriculum to the majority, the instructor may easily fall into the trap of mediocrity, which inevitably dissatisfies some students (Casey, Casiello, Gruca-Peal, and Johnson, 1995; Reid, Clunies-Ross, Goacher & Vile, 1981).

2. Departments and disciplines
   Social Science and Humanities tutors are rated more favourably than STEM subjects (Cashin, 1992; Centra, 1993).

3. Insufficient sample size

Both the size of the class and the response rate are important factors that should be taken into account when reviewing SET outcomes. To summarise, validity consists of threats from both external and internal validity. Evaluation University MEQ, the population external validity, external replication and one of internal validity may have threats to the validity. With regards to the reliability of University MEQ, bias from tutors, students and other variables may have threats to validity.
Discussion

Instrument design
MEQs typically comprise a mix of open and closed questions. This enables collection of standardised data whilst also allowing students to express details and information that might have not been considered by the designer (Jeffrey, et al, 2017).

The following are proposed for the design of the questionnaire:

- Purpose and goal of the questionnaire should be clearly stated before the MEQs are posted and during the response period so that students understand the value of participating.

- The questions asked impact the level of engagement so adequate consideration should be given to the design and focus of the questions. Adequate information should be provided with the questionnaire to clearly indicate the benefits of responding to the questionnaire for the students.

- Asking questions that require an extended response or at least a "content" answer.

- The questions so be designed in a way so that the task is clear to students. Questions such as "how satisfied are you…" do not often lead to productive answers. Consideration should also be given to students’ perception of the words such as “satisfied “that might have different meaning with different cultures.in designing the questions typical misconceptions that might lead students to provide a confused response. Moreover, language barriers ought to be considered in designing the questions.

The current version of the questionnaire includes seven questions. The low number of questions is designed so that the students to spend a short time and respond to the questionnairer, particularly as they are completing the survey for multiple modules. However, it appears that some questions may be confusing and the questionnaire fails to assess teaching quality way. Appendix1 outlines some of the potential confusion created by the existing MEQ which could be addressed through a redesign of the instrument.

Conclusions and Recommendations

To conclude, SET has three purposes: institutional, teaching and academic promotion. The needs of the stakeholder groups should be taken into account when addressing the design of the University MEQs to ensure that the intended purpose is achieved. The primary recommendation of this report is to undertake a staff-student partnership to agree the purpose of the MEQs and co-design a revised instrument that meets the stated purpose.

University MEQs should be used in conjunction with other quantitative evidence such as class average attendance rate, average, maximum and minimum marks as well as qualitative response analysis which would help build a more accurate overall picture of the class and help to reduce the bias effects outlined above. Further, full MEQ and other teaching related data should be provided to promotion panels to avoid the cherry picking of comments or data by applicants.
Where appropriately developed and deployed any MEQs have the potential to offer valuable insight into the student experience both within cohorts, across the institution and on a longitudinal basis.
Appendix 1 – Analysis of current MEQ questions

1. It was clear to me what I would learn and why
   **Aim:** To assess the clarity of the module content
   **Weakness:** the combination of “what, why, how” makes the responses invalid. There is no way for individuals to state if they have struggled with one or two aspects of the question since all three elements are attached under one question. It might be clear to individuals what they will learn but not clear how and when and

2. The materials on Canvas were useful and relevant
   **Aim:** To assess the validity and relevance of the module content on the VLE
   **Weakness:** It is not clear to which content this question refers to. Students might find the content relevant but not useful and vice-versa resulting in dissatisfaction with the questionnaire and invalid responses.

3. The recommended reading lists were appropriate and up to date
   **Aim:** To assess the reading list quality
   **Weakness:** The question mixes two aspects of the reading list and students have no means of choosing one to feedback on resulting in dissatisfaction with the questionnaire and invalid responses.

4. I have received feedback on my work (including queries after teaching sessions, in person, or by e-mail) which may have helped my understanding/clarified things I didn’t know, helped to explain a grade, or identified areas for improvement
   **Aim:** To assess feedback level
   **Weakness:** Mix of various elements of feedback in one question. The question refers to summative and formative feedback and can be confusing for students to respond to. There is no possibility for students to single out one aspect from the rest and to point out the challenges. A tutor might be very good with summative feedback on their module which support students to benchmark their learning at the end of the instructional unit. However, they may fail to offer adequate formative feedback and monitor students’ ongoing learning during the module.

5. Clear information about the assessment of this module, and the marking criteria, was provided
   **Aim:** To assess the assessment instruction and marking criteria
   **Weakness:** none
   **Suggestion:** to keep the question

6. Teaching accommodation and facilities were satisfactory
   **Aim:** To assess the learning environment
   **Weakness:** Tutors and students are assigned to certain teaching spaces and moving from one teaching space to another is usually extremely challenging due to limited spaces available. In addition, the question is not in line with other questions which focuses on students learning. Contact hours per week in physical teaching spaces are considerably
lower compared to self-study hours per week. Consequently, whilst the physical teaching spaces follow a minimum required standard which is looked after by the state, it might be more important to ask students how convenient and engaging was the VLE for them.

7. **Overall I was satisfied with the module**
   **Aim:** To Assess the students’ satisfaction
   **Weakness:** satisfaction might have a different meaning in different cultures. In addition, it might imply a different feeling with different individuals.
References


