
This is the accepted version.

©2015, Taylor & Francis

This is an Accepted Manuscript of an article published by Taylor & Francis in Australasian journal of engineering education in 2015, available at: http://www.tandfonline.com/doi/full/10.7158/D14-003.2015.20.1

Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30070163
The impact of increasing course enrolment on student evaluation of teaching in engineering education

S. Palmer†
Deakin University, Geelong, Victoria

W. Hall
Griffith University, Southport, Queensland

† Corresponding author Stuart Palmer can be contacted at spalm@deakin.edu.au.

ABSTRACT: Student evaluation of teaching (SET) is important, commonplace and may be used in staff performance management. The SET literature suggests that class size is a negative systematic influence on SET ratings. In this paper we investigate time-series SET data from a large first-year engineering class where a decline in SET ratings was observed over time as course enrolment increased. We observe a negative halo effect of increasing class size on mean SET ratings and conclude that increasing course enrolment leads to a significant reduction in all mean SET ratings, even when the course learning design remains essentially unchanged. We also find an additional differential effect of increasing course enrolment on mean SET ratings. We observe that the marginal reduction in mean SET ratings for each additional student in the course enrolment is greater for those aspects of the student learning experience that are likely to be most directly impacted by increasing class size. We provide implications for practice from these findings.
KEYWORDS: engineering education; large classes; student evaluation of teaching.

1 INTRODUCTION

Student evaluation of teaching (SET) has a long history, has grown in prevalence and importance over a period of decades, and is now common-place in many universities internationally (Davies et al., 2007; Denson et al., 2010; Hulpiau et al., 2007; Kember et al., 2002; Lemos et al., 2011; Narasimhan, 2001). SET data are collected for a range of purposes, including: as diagnostic feedback to improve the quality of teaching and learning; as an input to staff performance management processes and personnel decisions such as promotion for staff; to provide information to prospective students in their selection of courses and programs; and as a source of data for research on teaching (Kember et al., 2002; Marsh and Roche, 1993; Neumann, 2000).

Rovai et al. (2006) report that while SET research provides mixed results, there is evidence that, for course-related factors, smaller classes are rated more favourably than large classes, upper-year-level classes are rated more favourably than lower-year classes, and that there are rating differences between discipline areas. While additional course-related factors are also noted, other reviews of the literature on SET also identify these three factors as commonly reported systematic influences on SET ratings (Davies et al., 2007; Neumann, 2000; Palmer and Smith, 2013). In particular, research into student learning indicates a significant negative correlation between class size and quality of learning (Trigwell, 2011), and that class size is a key process variable in higher education that is predictive of student learning outcomes and satisfaction (Gibbs, 2010). A large investigation in a US college of Engineering found a range of systematic
influences on SET ratings, with class size being a particularly strong and significant negative influence - “Faculty [academic staff] assigned to teach large courses are likely to receive lower aggregate SET scores.” (Johnson et al., 2013, p. 311)

While the use of SET data may have originally been primarily for formative purposes to improve teaching and learning (Rovai et al., 2006), they are also increasingly used for summative judgements of teaching quality and teaching staff performance that have implications for personnel decision making (Johnson et al., 2013; Neumann, 2000; Simpson and Siguaw, 2000; Stratton et al., 1994). There may be an acceptance of the need for SET, however there remains no universal consensus as to what constitutes quality in university teaching and learning (Lemos et al., 2011; Walker, 2012), and the increasing use of SET for high-stakes decision making puts pressure on institutions to ensure that their SET practices are sound, equitable and defensible (Neumann, 2001). In particular, Johnson et al. (2013) note that, “Class size is important because faculty generally have little control over their class size; and if significant, the effects of class size should be included in interpretations of evaluations.” (Johnson et al., 2013, p.291)

If SET data are to be used effectively for the enhancement of teaching and learning quality, or in fact for any purpose, it is important that the rating results be communicated in a manner that is understandable and provides a sound basis for rational decision making (Lemos et al., 2011). Simply reporting mean SET ratings is problematic and potentially misleading for decision making; there is a need for an interpretation system that takes into consideration the known systematic influences on SET data (Rovai et al., 2006; Smith, 2008). One long-standing model of providing guidance on the interpretation of SET data is the Rating Interpretation Guides (RIGs) system (Lemos et al., 2011; Neumann, 2000; Palmer and Smith, 2013; Santhanam et
al., 2000). Although the specifics of various RIGs-style systems vary, the essential element is the provision of a norm-based set of benchmarks for the ranking or comparison of SET results. These benchmarks are based on a set of units of study that are similar in certain relevant respects—typically class size, class year level and discipline grouping—to the target course.

SET and related teaching quality data are often subjected to only superficial analyses that yield results of limited or questionable value—more explanatory and predictive utility could be derived from these data sources if more sophisticated analysis methods were employed (Gibbs, 2010). Principal Component Analysis (a dimensional reduction technique) was applied to a large university SET data collection to show a general negative association between student satisfaction and class size (Costantini et al., 2010). Based on the wider SET literature such a finding is not unexpected. However, additional analysis revealed a strong negative association between class size and ‘adequacy of lecture hall’, pointing to a more nuanced relationship between the negative systematic influence of class size and individual SET rating items that capture aspects of the student experience that are directly impacted by class size. For simplicity in comparison and/or performance evaluation, a single aggregate SET score is often derived by combining (often through simple unweighted averaging) the mean ratings obtained from all of the items on a SET survey (Johnson et al., 2013). However, such an approach cannot adequately represent the multidimensionality of teaching, and there is no automatic reason to assume that averaging mean scores from an assortment of SET items produces any meaningful information (Marsh and Roche, 1997). Marsh and Roche (1997) urge that the distinct components of SET be treated as the ‘apples and oranges’ that make up effective teaching, rather than treated as a single ‘puree’.
It was the observation of one of the authors in teaching a first-year course of study in engineering materials over several years that, as course enrolments rose significantly over time, mean ratings for the course had significantly decreased for all items on the institutional SET survey, even though no substantive changes had been made to the learning design of the course. This paper presents an investigation into the impact of increasing course enrolment on student evaluation of teaching ratings in a large first-year engineering class. It draws on SET data from a four year period, generated from a SET survey instrument with a common set of six student evaluation items over that time. Importantly, the key identified systematic influences on SET ratings of year level and discipline area are held constant, as are the course instructor and the course learning design, permitting a clear focus on the impact of increasing course enrolment. This investigation explores in a fine-grained way the relationships between course enrolment and the individual SET survey items, characterising the impact of class size of different aspects of the student learning experience.

2 CONTEXT – 1502ENG ENGINEERING MATERIALS

The programs offered by Griffith School of Engineering on the Gold Coast campus were restructured in 2010 to facilitate a common first-year for all its specialist majors (Hall et al., 2012). Since the introduction of the common first-year, there has been some amendment to the five specialist majors offered on the Gold Coast campus, but the common first-year has remained a feature. The current programme offerings on the Gold Coast are:

- Civil Engineering;
- Mechanical Engineering;
- Electrical and Electronic Engineering;
• Mechatronic Engineering; and
• Electronic and Biomedical Engineering.

With the introduction of the common first-year, a new course ‘1502ENG Engineering Materials was created. This course, 1502ENG, still exists in the five specialist majors. It aims to provide an introduction to materials science and technology, and other than the significant year-on-year increases in student enrolment (from an initial 212 students to 357 students in 2013), has remained largely unchanged since its introduction.

The course focuses on a traditional delivery mode with two hours per week of lectures and a one hour tutorial. Initially, three laboratory practical sessions were included in the course, but this was later revised to two following student feedback. This has been the only ‘structural’ change to the course. Assessment for the initial offering was via a mid-semester test (15%), the three laboratory practical tasks (5%, 10% and 10%), a group assessment (10%) and a final examination (50%). In the second and subsequent offerings of the course, the removal of the third laboratory required a redistribution of assessment marks. The marks for the third laboratory (10%) were reallocated to the examination which therefore rose to 60%. In 2012 and 2013, to provide additional feedback to students, the mid-semester test was assessed solely via a multiple choice test, and was followed by a feedback/debriefing session that included working through the mid-semester test content. At the same time the final examination was also changed from a multiple choice and written assessment to solely focus on multiple choice questions of various weightings (marks). Table 1 presents a summary of the assessment items and the percentage allocation of marks to each item for the four offerings of 1502ENG so far (initial 2010 offering, 2011, 2012 and 2013).

Table 1: Summary of assessment items for 1502ENG 2010 – 2013.
<table>
<thead>
<tr>
<th>1502ENG Assessment item</th>
<th>Marks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2010</td>
</tr>
<tr>
<td>Lab 1</td>
<td>5</td>
</tr>
<tr>
<td>Lab 2</td>
<td>10</td>
</tr>
<tr>
<td>Lab 3</td>
<td>10</td>
</tr>
<tr>
<td>Mid-semester test</td>
<td>15*</td>
</tr>
<tr>
<td>Group Assessment (written report)</td>
<td>10</td>
</tr>
<tr>
<td>Examination</td>
<td>50*</td>
</tr>
</tbody>
</table>

* Multiple choice and written questions  # Multiple choice questions only

Griffith University employs a ‘Student Experience of Course’ (SEC – formerly ‘Student Evaluation of Course’) online survey as part of its student evaluation of teaching, quality improvement and staff performance management processes. Note that Griffith University also periodically uses a ‘Student Experience of Teaching’ survey, which is locally abbreviated as ‘SET’, but is not used as part of the analysis presented here, and should not be confused with our use of the acronym SET here (as the widely used abbreviation for ‘student evaluation of teaching’). The SEC survey is opened for students to voluntarily complete towards the end of the teaching period, but prior to the examination period, so SEC data are unlikely to overly influenced by students’ perceptions of how they performed on the exam and/or the course overall. During the period under consideration here, the SEC survey contained six common items for students to respond to. Each item is presented as a statement, to which students indicate their level agreement on a five point scale of the form: Strongly agree (5); Agree (4); Neutral (3); Disagree (2); and Strongly disagree (1). The six SEC items are:

- SEC1 - This course was well-organised.
- SEC2 - The assessment was clear and fair.
• SEC3 - I received helpful feedback on my assessment work.
• SEC4 - This course engaged me in learning.
• SEC5 - The teaching (lecturers, tutors, online etc) on this course was effective in helping me to learn.
• SEC6 - Overall I am satisfied with the quality of this course.

During the period under consideration, the course convenor remained the same and all moderation of assessment results was undertaken by the course convenor. As the class size increased there was a need for more tutorial and laboratory classes, and this necessitated a more significant tutor (PhD student) involvement. There was therefore a notable change in tutorial support, but the academic management of the course remained consistent during the period under consideration. Moreover, unlike the SET survey instrument which focusses primarily on the performance and characteristics of individual academic staff members involved in teaching, the SEC survey instrument used here focusses primarily on the characteristics of the course and not the teaching staff.

Following the completion of the SEC survey period, a ‘SEC Detailed Report’ is provided to the course convenor which tabulates the student response data. The SEC Detailed Report contains no information capable of identifying any student respondent. For each of the six SEC items, the SEC Detailed Report includes the following data:
• the distribution of individual SEC rating scores;
• the mean rating;
• the standard deviation of the mean rating;
• the median rating; and
3 METHODOLOGY

Approval was sought from the Griffith University Human Research Ethics Committee to use the SET data presented on the SEC Detailed Reports for the course 1502ENG over the period 2010 to 2013, and approval was granted. The SEC data were tabulated and charted to examine any trends in mean ratings for each SEC item with increasing course enrolment. Independent two-sample t-tests were performed on the data sets for each SEC item to test if the mean ratings obtained for 2010 and 2013 were significantly different. Simple linear regression was employed to characterise the overall trends in mean SEC ratings. Confidence intervals were computed for the simple linear regression scalar multiples obtained for all SEC items to identify the presence of any groupings in the relationship between mean SEC ratings and course enrolment. The results obtained and a discussion of the observed results are presented.

We acknowledge that the underlying SET ratings provided by students are derived from response scales and are fundamentally ordinal in nature. However, students are generally aware that the data, for practical purposes, are treated as originating from a five point interval scale, and are reported and used via the SEC system as a mean rating out of five. The use of ordinal data in many parametric statistical procedures, while commonplace in the social sciences, is not universally accepted as valid. However, there is a significant body of research that has demonstrated the practical utility of analysis of ordinal data, based on the robustness of many parametric methods to significant departures from assumptions about the underlying data,
including departures from normality and ‘intervalness’ that might be present in ordinal scale data (Jaccard and Wan, 1996; Norman, 2010). We further acknowledge that we are drawing on a single source of time series data (SEC ratings), and that it would be desirable to be able to triangulate and/or supplement any findings with complementary data sources such as student demographics or qualitative evaluation data from students and staff, and/or comparisons with results from similar units of study at other universities. The access to, and use of, such additional data sources were beyond the scope of the current investigation.

4 RESULTS

For each year from 2010 to 2013, Table 2 shows the following data received on the SEC Detailed Report for the course 1502ENG:

- course enrolment;
- SEC response rate;
- mean rating for each SEC item; and
- the standard deviation of each mean rating in parentheses.

Table 2: Summary of SEC rating data 2010 – 2013.

<table>
<thead>
<tr>
<th>Year</th>
<th>Course enrolment</th>
<th>Response rate</th>
<th>SEC1</th>
<th>SEC2</th>
<th>SEC3</th>
<th>SEC4</th>
<th>SEC5</th>
<th>SEC6</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>212</td>
<td>58.5%</td>
<td>4.77 (0.44)</td>
<td>4.52 (0.59)</td>
<td>4.18 (0.83)</td>
<td>4.47 (0.63)</td>
<td>4.72 (0.52)</td>
<td>4.63 (0.53)</td>
</tr>
<tr>
<td>2011</td>
<td>236</td>
<td>53.8%</td>
<td>4.57 (0.62)</td>
<td>4.23 (0.66)</td>
<td>4.02 (0.84)</td>
<td>4.23 (0.76)</td>
<td>4.54 (0.70)</td>
<td>4.34 (0.72)</td>
</tr>
<tr>
<td>Year</td>
<td>Enrolments</td>
<td>Response Rate</td>
<td>SEC1</td>
<td>SEC2</td>
<td>SEC3</td>
<td>SEC4</td>
<td>SEC5</td>
<td>SEC6</td>
</tr>
<tr>
<td>------</td>
<td>------------</td>
<td>---------------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>2012</td>
<td>287</td>
<td>23.7%</td>
<td>4.66</td>
<td>4.25</td>
<td>3.93</td>
<td>4.32</td>
<td>4.35</td>
<td>4.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.70)</td>
<td>(1.01)</td>
<td>(1.23)</td>
<td>(0.87)</td>
<td>(0.93)</td>
<td>(1.01)</td>
</tr>
<tr>
<td>2013</td>
<td>357</td>
<td>35.6%</td>
<td>4.20</td>
<td>3.96</td>
<td>3.42</td>
<td>3.96</td>
<td>3.99</td>
<td>3.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.71)</td>
<td>(0.92)</td>
<td>(1.06)</td>
<td>(0.78)</td>
<td>(0.92)</td>
<td>(0.80)</td>
</tr>
</tbody>
</table>

The course enrolments and obtained response rates mean that the respondent samples obtained in 2010, 2011 and 2013 are almost identical in size. The respondent sample obtained in 2012 was somewhat smaller, however the parametric statistical testing employed takes into account the size and spread of the various respondent samples when estimating the significance level of the observed test results – a smaller sample naturally increases the level of uncertainty associated with a test result. We report the measures of uncertainty associated with all statistical testing performed. Figure 1 shows the mean rating for each SEC item versus course enrolment – note that both axes have expanded scales.

**Figure 1:** Mean SEC ratings versus course enrolment.
In Figure 1 it can be seen that the mean ratings for three of the SEC items decrease monotonically with increasing course enrolment (SEC3, SEC5 and SEC6) and three do not (SEC1, SEC2 and SEC4). However, there is a general downward trend in mean SEC ratings across the time period, and all SEC items are lower in 2013 than they were in 2010. Independent two-sample $t$-tests were performed on the data sets for each SEC item to test if the mean ratings obtained for 2010 and 2013 were significantly different. The first step in this analysis is to check the homogeneity of variance in the sample data, to determine the most appropriate form of the $t$-test to apply. The variance of the 2010 and 2013 data sets were not significantly different for SEC2, SEC4 and SEC6, and for these three items the standard $t$-test was used. For SEC1, SEC3 and SEC5, the variance of the 2010 and 2013 data sets were significantly different, and for these three items Welch's $t$-test, which is robust to unequal variance, was used. Table 3 shows the results of the $t$-tests, including the $t$ statistic, the applicable degrees of freedom and the resultant $p$ value for each SEC comparison.

<table>
<thead>
<tr>
<th>Test parameter</th>
<th>SEC1</th>
<th>SEC2</th>
<th>SEC3</th>
<th>SEC4</th>
<th>SEC5</th>
<th>SEC6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>7.57</td>
<td>5.67</td>
<td>6.35</td>
<td>5.65</td>
<td>7.71</td>
<td>8.28</td>
</tr>
<tr>
<td>DF</td>
<td>213.0</td>
<td>249</td>
<td>237.7</td>
<td>249</td>
<td>199.6</td>
<td>248</td>
</tr>
<tr>
<td>$p$</td>
<td>1.24x10^{-12}</td>
<td>3.88x10^{-8}</td>
<td>1.07x10^{-9}</td>
<td>4.28x10^{-8}</td>
<td>6.83x10^{-13}</td>
<td>1.08x10^{-13}</td>
</tr>
</tbody>
</table>

Generally, the SEC data sets were not normally distributed, but the independent two-sample $t$-test is relatively robust to departures from normality in the data, and the very small $p$ value significance results obtained show unequivocally that the mean SEC ratings for 2013 are all
significantly lower than those obtained in 2010. In Figure 1 the various mean SEC ratings are shown as data points for each SEC item for each course enrolment level. As mean ratings, each data point in Figure 1 summarises a distribution of ratings obtained from students for each SEC item each year over the period 2010 to 2013. The standard deviation values given for each mean SEC rating in Table 2 provide additional summary information about the distribution of ratings for each SEC item in each period. By way of illustration, Figure 2 shows the mean SEC rating trajectory for SEC2 (as shown in Figure 1) over the top of the distribution of individual SEC rating scores obtained for SEC2 each year/course enrolment – note that the course enrolment axis has an expanded scale.

Figure 2: Detailed SEC rating data and regression model for SEC2.

The individual SEC rating counts are shown by the ‘bubbles’ in Figure 2, where the bubble area is proportional to the SEC rating count, and the actual numerical SEC ratings counts are given adjacent to each bubble. For example, in 2011 (when the course enrolment was 236), the rating distribution for SEC2 was: nil ratings of one; 1 rating of two; 13 ratings of three; 69 ratings of
four; and 44 ratings of five. The rating count distribution may be informative and important in more detailed analysis, however it is generally the summary mean ratings that are used in reporting and staff performance evaluations.

In Figure 1 it can be seen that the various mean SEC rating trajectories follow an overall trend downward with increasing course enrolment, suggesting a negative relationship between course enrolment and mean SEC ratings. One method for characterising the overall trend in the mean SEC ratings, while at the same time taking into account the detailed trajectory of the observed mean SEC ratings and the variance/dispersion in the underlying rating count distributions, is to assume that the relationship between mean SEC ratings and course enrolment is linear and perform a simple linear regression analysis. Simple linear regression yields a model for estimating the mean SEC rating at any level of course enrolment as a scalar multiple of course enrolment plus a constant offset. The scalar multiple is the slope of the modelled/predicted linear relationship between course enrolment and mean SEC rating. The regression model is the result of minimising the sum of the square of the errors between the observed actual dataset and the responses predicted by the linear approximation.

Simple linear regression assumes that observed values for mean SEC rating at different course enrolments are normally distributed, and that the variance in the observed values of mean SEC ratings is constant across different course enrolments. Figure 2 suggests that these assumptions do not strictly hold for SEC2, and investigation shows that they do not hold for any of the other SEC items. However, the relative difference in variance at different course enrolments is comparatively small for all SEC items, and the same general trend in SEC rating distributions with increasing course enrolment (broadening and shifting lower due to the presence of more lower ratings) is observed in all SEC items, suggesting that the linear regression models
obtained for the SEC items can be validly compared. The linear trajectory based on predicted values for SEC2 obtained from the regression model is also shown in Figure 2. Figure 3 shows the simple linear regression scalar multiples obtained for all SEC items, and their estimated 90 per cent confidence intervals.

![Simple linear regression scalar multiples obtained for all SEC items.](image)

**Figure 3:** Simple linear regression scalar multiples obtained for all SEC items.

The regression scalar multiple is not simply the slope of a straight line between the 2010 and 2013 mean ratings for a SEC item, rather it takes into account the detailed trajectory of all the observed mean ratings for a SEC item. The scalar multiple in the regression model is an estimate of the marginal reduction in mean SEC rating for each additional student in the course enrolment, and provides a measure of the sensitivity of each SEC item to increasing course enrolment. Because the regression models here are based on available course enrolment data over the range 212 to 357 only, caution should be taken in extrapolating the models significantly beyond that range.
5 DISCUSSION

As apparent in Table 2 and Figure 1, the mean ratings for all SEC items trend downward with increasing course enrolment over time. And, as Table 3 shows, the 2013 mean ratings for all SEC items are significantly lower than for 2010. Simple linear regression was employed to characterise the overall trend in mean SEC ratings, taking into account the detailed trajectory of the observed mean SEC ratings and the variance/dispersion in the underlying rating count distribution. The locations of the 90 per cent confidence intervals shown in Figure 3 suggest that there are two distinct groups of linear relationship between mean SEC rating and course enrolment – one with a regression multiple of approximately -0.0032 (comprised of SEC1, SEC2 and SEC4), and another with a regression multiple close to 50 per cent larger at approximately -0.0047 (comprised of SEC3, SEC5 and SEC6).

The first group with the smaller negative multiple includes the SEC items:

- SEC1 - This course was well-organised.
- SEC2 - The assessment was clear and fair.
- SEC4 - This course engaged me in learning.

SEC items 1 and 2 clearly relate to organisational aspects of the course that, while highly desirable to be done well, are not typically impacted significantly by course enrolment. Large course enrolments may necessitate competing for more and/or larger classrooms, which might in turn lead to less desirable timetabling from a student perspective, but most course organisation issues apply equally to classes large and small. Group-based assessment that might be employed in response to large enrolments has been known to raise issues regarding equitable contributions between group members, but generally the clarity and fairness of assessment is
independent of class size. Student engagement has been defined as, “the quality of effort students themselves devote to educationally purposeful activities that contribute directly to desired outcomes” (Hu and Kuh, 2002, p. 555). While the effort that students make might be influenced by some aspects of their learning environment, including class size, it is also significantly influenced by personal factors that are innate and/or contextual to individual students, and not related to the course enrolment.

The second group with the larger negative multiple includes the SEC items:

- SEC3 - I received helpful feedback on my assessment work.
- SEC5 - The teaching (lecturers, tutors, online etc) on this course was effective in helping me to learn.
- SEC6 - Overall I am satisfied with the quality of this course.

Assessment costs are largely proportional to student numbers (Gibbs, 2010), and, expanding student numbers, reducing real funding per student and workload models that don’t properly recognise class size may all limit the practical ability of staff to provide timely, relevant and personalised feedback to individual students (Walker, 2012). Class size may influence the teaching approach used by a teacher and/or impact on the amount of personal communication or attention that a teacher can give to any particular student (Adams et al., 1996; Centra and Gaubatz, 2000), and large numbers of students may mean that close contact with academic staff outside of class and/or access to remedial assistance may be more limited (Gibbs, 2010). There is strong evidence that students’ overall ratings of large classes reflect a systematic, general downward bias as class size increases that is largely independent of the course learning design and the teacher (Gibbs, 2010; Johnson et al., 2013).
The findings here suggest that increasing course enrolment has a general negative influence on mean student SET ratings, but that this influence is not uniform. We observe that the marginal reduction in mean SEC rating for each additional student in the course enrolment is greater for those SEC items relating to aspects of the student learning experience that are likely to be most materially and directly impacted by increasing class size. The interpretation of SET data, particularly in relation to staff performance management, needs to be cognisant of these observed systematic influences (both common mode and differential) on mean SET ratings.

It is apocryphally stated that the best way to improve large classes is to have smaller classes. While it is beyond the scope of this paper to provide a detailed investigation of strategies for improving large class learning and teaching, there are many documented examples from across a range of disciplines that do suggest means to improve students’ perceptions of, engagement with, and ultimately learning outcomes from, large enrolment classes. Suggested activities include pre-class reading followed with student group discussions, small-group active learning tasks, in-class instructor feedback, in-class quizzes and the use of audience response system (clicker) activities (Deslauriers et al., 2011); peer assessment in tutorials (Snowball and Sayigh, 2007); and peer assessment that provides not only timely formative feedback, but is also combined with elements of assessment by staff to contribute to the overall student summative assessment (Wanous et al., 2009). It should be noted that these approaches are not simple, unproblematic solutions to all of the issues related to large classes – all of the cited examples note requirements for staff (and in some cases student) training and additional costs for equipment and/or staff time.
This paper presents an investigation into the impact of increasing course enrolment on student evaluation of teaching (SET) ratings in a large first-year engineering class. As is often reported in the literature, we observe a negative common mode or halo effect of increasing class size on mean SET ratings. We conclude that increasing course enrolment leads to a statistically significant reduction in all mean SET ratings, even when the course learning design remains essentially unchanged. Beyond this perhaps expected observation, we also find an additional differential effect of increasing course enrolment on mean SET ratings. Our regression analysis suggests that, while increasing course enrolment has a general negative influence on mean student SET ratings, this influence is not uniform. For the course enrolment range under consideration, we observe that the marginal reduction in mean SET ratings for each additional student in the course enrolment is greater for those SET items relating to aspects of the student learning experience that are likely to be most directly impacted by increasing class size. While there exists in the research literature the call for more sophisticated analysis of SET data, and the observation of nuanced relationships between class size and different aspects of SET data, we believe that the finding here of a quantifiable differential effect on SET ratings under varying course enrolment is a new contribution.

These findings present two implications for practice. Firstly, it is likely that a significant increase in class size will lead to a significant and systematic decrease in mean SET ratings generally, even if all other factors are held constant. Learning and teaching evaluation systems, and staff performance management systems, should take this factor into account, and a Rating Interpretation Guides (RIGs) style system is one mechanism for achieving this in a rational, equitable and defensible way. Secondly, actions taken to improve the student experience in large classes are likely to have the most impact when they are directed toward those aspects of
learning and teaching that students experience as individuals, for example, the quantum, quality and timeliness of individual feedback received on assessment items, and the amount of personal contact that individual students can have with teaching staff. In addition, we offer the methodology presented here as a useful basis for future research into the factors that systematically influence SET data.

REFERENCES


