The systematic influences on student evaluation of teaching in engineering education

Stuart Palmer

Faculty of Science, Engineering and Built Environment, Deakin University, Victoria, Australia
Corresponding Author's Email: spalm@deakin.edu.au

BACKGROUND
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. 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. 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. The School of Engineering at Deakin University in Australia offers undergraduate and postgraduate engineering programs, and these programs are delivered in both on-campus and off-campus modes.

PURPOSE
The paper presents a quantitative investigation of SET data for the School of Engineering at Deakin University to identify whether the commonly reported systematic influences on SET ratings of class size and year level are also observed here. The influence of online mode of offer is also explored.

DESIGN/METHOD
Deakin University's Student Evaluation of Teaching and Units (SETU) questionnaire is administered to students enrolled in every unit of study every time that unit is offered, unless it is specially exempted. Following data collation, summary results are reported via a public website. The publicly available SETU data for all School of Engineering units of study were collected for a two year period. The collected data were subjected to analysis of variance (ANOVA) analysis to identify any significant systematic influences on mean student SETU ratings.

RESULTS
SETU data from 100 separate units of study over the two year period were collected, representing 3375 sets of SETU ratings, and covering unit enrolment sizes from 12 to 462 students. Although this was a modest sized investigation, significantly higher mean ratings for some SETU items were observed for units with small enrolments, for postgraduate level units compared to undergraduate level units, and for units offered in conventional mode compared to online mode of offer. The presence of the commonly observed systematic influences on SET ratings was confirmed.

CONCLUSIONS
While the use of SET data may have originally been primarily for formative purposes to improve teaching and learning, they are also increasingly used for summative judgements of teaching quality and teaching staff performance that have implications for personnel decision making. 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, 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.

KEYWORDS
Student evaluation of teaching; Class size; Year level.

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, Hirschberg, Lye, Johnston, & McDonald, 2007; Denson, Loveday, & Dalton, 2010; Kember, Leung, & Kwan, 2002). 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 & 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). In the context of engineering education, a large investigation in a US college of Engineering found a range of systematic influences on SET ratings, including class size and class level (junior, senior, graduate) (Johnson, Narayanan, & Sawaya, 2013).

In more recent times, there has been dramatic growth in online education internationally that continues to this day (Allen & Seaman, 2010; Loveland, 2007; Mayadas, Bourne, & Bacsich, 2009). Rovai et al. (2006) note that much of the published research on SET relates to traditional classroom settings, and in an analysis of qualitative SET data (open-ended student written comments) they found a significant difference between the responses of students completing a wholly online version of a course compared to students completing an on-campus version of the same course – online students gave a more negative rating (Rovai, Ponton, Derrick, & Davis, 2006). Loveland (2007) notes the lack of research on the use of SET to evaluate online teaching, provides a study indicating significantly lower SET ratings for online classes compared to on-campus classes, and calls for additional research in this area. An institution-wide analysis of SET data at Deakin University confirmed the significant influence of class size, class level and discipline area on at least some SET ratings, and also found online mode of offer to significantly influence at least some SET ratings (Palmer & Smith, 2013).

This paper presents an investigation into the systematic influences on SET ratings for engineering programs at Deakin University. This investigation draws on publicly available SET data over a two year period – 100 units of study in total. Specifically, this investigation seeks to identify any observable significant systematic influence on SET ratings assignable to class size and/or class level and/or online mode of offer. Being limited to engineering programs, the commonly identified systematic influence on SET ratings of discipline area is usefully held constant. The relatively small scale of the investigation means that, rather than being definitive, the findings are an indicator for future work to further characterise the impact of any observed systematic influences on SET ratings, and for the development of methods to most effectively use collected SET data in the improvement of learning and teaching in engineering.

Deakin University context

The School of Engineering at Deakin University in Australia offers undergraduate and postgraduate engineering programs. These programs are delivered in both on-campus and off-campus modes. Off-campus students are typically mature aged, working full-time, have significant experience in an engineering-related job role, and may live remotely from the university campus, including overseas. In the past, off-campus students typically received a

substantial package of printed study materials that may have been supplemented with other learning resources such as video on tape or disk. More recently, virtually all off-campus learning resources have been migrated online, in the form of web pages, downloadable documents and streaming video. Off-campus students attend on-campus for mandatory minimum periods as required by engineering program accreditation, and on-campus students typically have access to the same online learning materials as off-campus students.

Deakin University has an 'Evaluation of Teaching and Units' procedure that requires Deakin University's Student Evaluation of Teaching and Units (SETU) questionnaire to be administered to students enrolled in every unit of study every time that unit is offered, unless it is specially exempted. This means that a large volume of SET data is collected annually at Deakin University. The SETU instrument, as a standardised, centrally administered questionnaire, currently consists of ten core items:

1. This unit was well taught.
2. The course materials in this unit were of high quality.
3. The workload in this unit was manageable.
4. Requirements for completing the assessment tasks in this unit were clear.
5. The teaching staff gave me helpful feedback.
6. The library resources met my needs for this unit.
7. I would recommend this unit to other students.
8. The technologies used to deliver the online content in this unit performed satisfactorily.
9. The on-line teaching and resources in this unit enhanced my learning experience.
10. This unit challenged me to learn.

SETU respondents rate each core item on a five point scale (1=strongly disagree; 2=disagree; 3=neutral; 4=agree; 5=strongly agree) with a 'not applicable' option included. In addition to the ten core items, SETU includes a rating of, and open-ended comment on, the quality of teaching for each nominated staff member involved in teaching on the unit, and open-ended comments on the 'best aspects' of, and 'aspects most in need of improvement' for the unit.

The SETU survey is open for a period of approximately one month at the end of each academic teaching period, and following collation of results, SETU data are reported via a public web site. Anyone with an interest can query the summary results for the ten core SETU items, based on a selection of teaching/evaluation period, faculty, school, unit and student enrolment location (Deakin University, 2014). The data reported for a unit include total enrolment, number of responses and computed response rate for the enrolment location(s) selected, and, number of responses, mean rating and standard deviation of the mean rating for each of the ten core SETU items. SETU results are publicly reported for a unit unless the number of responses is less than ten; the presumption being that anything less than ten responses is an unrepresentative sample size.

Method

Mean ratings for the ten core SETU items for all units belonging to the Deakin University School of Engineering and reporting data via the Deakin University SETU web site were collected for the six main academic teaching periods in 2011 and 2012. It is acknowledged that the underlying SET ratings provided by students are derived from response scales and are fundamentally ordinal in nature. However, students and staff are generally aware that the data, for practical purposes, is treated as originating from a five point interval scale, and are reported and used via the SETU 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 statistical 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 & Wan, 1996; Norman, 2010). Additionally, while the underlying SET ratings provided by students are ordinal data, the primary data element in this investigation is the mean of the ratings provided by students enrolled in a unit of study. The statistical analysis that follows is based on these unit-level mean ratings, where the central limit theorem suggests that, regardless of the distribution of the original data, the distribution of the means will tend towards normality (Aron, Aron, & Coups, 2009). With these methodological limitations acknowledged, the following analyses were performed.

For each of class size, class level and online mode of offer, appropriate groupings were identified, as outlined below. For each of these three factors a one-way analysis of variance (ANOVA) was undertaken to identify any significant systematic variation in the mean rating between factor groupings for the ten SETU items. Where Levene's test of homogeneity of variance failed, a robust ANOVA test using the Welch test statistic was performed instead. Where a significant difference was observed and there were more than two groupings for a factor (i.e., class size and class level), post-hoc group pair-wise comparisons were performed to assess which factor groupings had significantly different mean SET ratings. For the post-hoc tests, where equal variance could be assumed Tukey's 'honest significant difference' test was used, and where equal variance could not be assumed Tamhane's T2 test was used.

The proxy measure for class size was taken to be the officially recorded unit enrolment; acknowledging that this was the nominal class size, which may vary depending on actual class attendances for on-campus students, and that 'class size' may affect the experience of off-campus students in different ways to on-campus students. For the purpose of practical utility, class sizes are typically divided into bands/categories (small, medium, large, etc.) which can then be used later to interpret SET ratings based on class size (Neumann, 2000; Smith, 2008). Observed class size bands used in SET data analyses vary - <20, 20-74, >74 (Hippensteel & Martin, 2005); 20-39, 40-49, 50-69,70-119, 120-149, 160-175 (Ragan & Walia, 2010); 1-20, 21-40, 41-60, 61-80, 81-100, 101-150, 151-200 (Johnson et al., 2013). For the 100 units included in the analysis here, the enrolment ranged from 12 to 462. Four groups were chosen for the analysis presented here. ANOVA analysis is most robust when there are approximately equal numbers of cases in each group. Rank ordering the cases by unit enrolment and dividing into four equal groups gives class size groupings of 12-34, 35-56, 58-104 and 115-462.

Likewise, for class level grouping, many combinations can be observed in the SET literature - first-year, second-year, senior-year (third to fifth) and postgraduate (Neumann, 2000); first-year, second-year, and, third-year and above (Smith, 2008); first-year, second-year, third-year, and, fourth-year and higher (Santhanam & Hicks, 2002). Here three groups were chosen – first- and second-year, third- and fourth-year, and postgraduate.

A 'normal' unit of study in the engineering program at Deakin University includes students enrolled in both on-campus and off-campus modes of study. Historically, a distinction was made between units offered in off-campus mode, where students would be sent print-based learning materials, and wholly online mode, where students would receive all learning materials online. As noted above, off-campus students now effectively study in online mode, but for some of the period under consideration, some units were still nominally recorded as wholly online mode of offer. For the purposes of this investigation, those units tagged as offered in wholly online mode and those offered in off-campus mode exclusively (no on-campus students enrolled) are collectively considered as being offered in online mode. All other units have a predominately on-campus enrolment, but may contain a proportion of off-campus students. This enrolment mode grouping schema is best considered as conventional mode of offer versus online mode of offer.

In all following analyses, the significance level used was $p < 0.02$. This level of statistical significance indicates that you could expect to see the observed differences in group means only one time in 50 if similar samples were drawn at random from their parent populations.

That is, if a significance level of $p < 0.02$ is observed, it is unlikely that the sample group means are an artefact of chance, and that they actually represent groupings with significantly different distributions of SETU ratings. The term SET is used to denote student evaluation of teaching generically. The term SETU is used to refer specifically to the student evaluation of teaching and units SET instrument employed at Deakin University. A discussion of the observed results is also presented.

Results and Discussion

The data extracted from the SETU reporting web site and used in the analysis here included mean rating sets for 100 units of study and represented 3375 sets of SETU ratings. The data set includes responses from student enrolled in all modes of study – on-campus, off-campus, and wholly online.

For each of the ten SETU items, a one-way ANOVA was attempted for the mean rating as the dependent variable against class size grouping. A number of significant differences were observed, and post-hoc testing revealed that the significant differences were primarily between the <35 group and all other groups. Based on this result, the three groups with enrolments greater than 34 were pooled to form a >34 group, and the one-way ANOVA test was repeated for the two groups <35 and >34. For all ten SETU items in the two group ANOVA test, Levene's test of homogeneity of variance passed, so a standard ANOVA test was performed. A significant difference in mean SETU rating between the two class size groupings was observed for all but two SETU items. Table 1 presents a summary of the statistical test results and a measure of effect size based on eta squared (η^2).

Table 1: Relationship between mean SETU rating and class size grouping

SETU item	Test statistic	Significance	Effect size
1. This unit was well taught	**$F = 7.45$**	**$p < 0.008$**	**$\eta^2 = 0.071$**
2. The course materials were of high quality	**$F = 8.23$**	**$p < 0.006$**	**$\eta^2 = 0.078$**
3. The workload in this unit was manageable	**$F = 6.45$**	**$p < 0.013$**	**$\eta^2 = 0.062$**
4. Assessment tasks in this unit were clear	**$F = 6.97$**	**$p < 0.010$**	**$\eta^2 = 0.066$**
5. The teaching staff gave me helpful feedback	**$F = 7.25$**	**$p < 0.009$**	**$\eta^2 = 0.069$**
6. The library resources met my needs for this unit.	$F = 3.69$	$p > 0.057$	$\eta^2 = 0.036$
7. I would recommend unit to other students	**$F = 13.85$**	**$p < 0.001$**	**$\eta^2 = 0.124$**
8. Technologies used performed satisfactorily	**$F = 7.93$**	**$p < 0.006$**	**$\eta^2 = 0.075$**
9. On-line enhanced my learning experience	**$F = 7.41$**	**$p < 0.008$**	**$\eta^2 = 0.070$**
10. This unit challenged me to learn.	$F = 1.04$	$p > 0.309$	$\eta^2 = 0.010$

For all ten SETU items the <35 class size group had higher mean ratings, and for the eight items **in bold** in Table 1 this difference was significant at the $p < 0.02$ level. The average eta squared for all ten items was 0.066 – that is, on average, the observed difference in mean SETU ratings between the two class size groupings was 6.6 per cent. However this ranged up to 12.4 per cent for SETU item 7 (I would recommend this unit to other students), which is important as this item is often taken as an overall SETU proxy indicator. It has been posited that 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, Neumann, & Rytmeister, 1996; Centra & Gaubatz, 2000), both of which may negatively impact on student perceptions of teaching as class size increases. Here the observed impact of class size, between the two groupings used, seems to be particularly pervasive, with a halo effect extending to some aspects of unit learning design that would

seem to be relatively independent of class size, i.e., clarity of assessment requirements and performance of online technologies. 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).

As above, for each of the SETU items, a one-way ANOVA was attempted for the mean rating as the dependent variable against class level grouping. Only one significant difference was observed (SETU item 7), and post-hoc testing revealed that the significant difference was between the postgraduate group and the two undergraduate groups. Based on this result, the two undergraduate groups were pooled, and the one-way ANOVA test was repeated for the two groups undergraduate and postgraduate. Levene's test of homogeneity of variance passed, so a standard ANOVA test was performed. A significant difference in mean SETU rating between the two class level groupings was observed for SETU 7. Table 2 presents a summary of the statistical test results and a measure of effect size based on eta squared (η^2).

Table 2: Relationship between mean SETU rating and class level grouping

SETU item	Test statistic	Significance	Effect size
1. This unit was well taught	$F = 2.27$	$p > 0.134$	$\eta^2 = 0.023$
2. The course materials were of high quality	$F = 2.81$	$p > 0.096$	$\eta^2 = 0.028$
3. The workload in this unit was manageable	$F = 0.12$	$p > 0.721$	$\eta^2 = 0.001$
4. Assessment tasks in this unit were clear	$F = 0.88$	$p > 0.349$	$\eta^2 = 0.009$
5. The teaching staff gave me helpful feedback	$F = 2.73$	$p > 0.101$	$\eta^2 = 0.027$
6. The library resources met my needs for this unit.	$F = 0.00$	$p > 0.960$	$\eta^2 = 0.000$
7. I would recommend unit to other students	**$F = 5.84$**	**$p < 0.018$**	**$\eta^2 = 0.056$**
8. Technologies used performed satisfactorily	$F = 0.74$	$p > 0.392$	$\eta^2 = 0.007$
9. On-line enhanced my learning experience	$F = 1.27$	$p > 0.261$	$\eta^2 = 0.013$
10. This unit challenged me to learn.	$F = 0.72$	$p > 0.397$	$\eta^2 = 0.007$

Here a significantly higher mean rating from postgraduate students (compared to undergraduate students) was observed only for SETU item 7 (**in bold**), but as noted previously, this item is often taken as an overall SETU proxy indicator, so still a potentially important result. The effect size here is relatively small – the observed difference in mean ratings between the two class level groupings for SETU item 7 was 5.6 per cent. Johnson et al. (2013) note other research indicating that postgraduate students gave higher SET ratings, and based on their own research in an engineering education context, observed a significant correlation between course level and mean SET rating. It has been suggested that this commonly observed influence on SET ratings is related to student maturity, and that after several years at university, older students have more realistic expectations of their university experience (Denson et al., 2010) or, at least in some discipline areas, students become more independent in their learning in the later years of their study and/or beyond undergraduate study (Adams et al., 1996).

As above, for each of the SETU items, a one-way ANOVA was attempted for the mean rating as the dependent variable against online mode of offer. Only one significant difference was observed (SETU item 10) Levene's test of homogeneity of variance passed, and because online mode offer contains only two groups, no post-hoc testing was required. Table 3 presents a summary of the statistical test results and a measure of effect size based on eta squared (η^2).

Table 3: Relationship between mean SETU rating and online mode of offer grouping

SETU item	Test statistic	Significance	Effect size
1. This unit was well taught	$F = 2.32$	$p > 0.130$	$\eta^2 = 0.023$
2. The course materials were of high quality	$F = 0.58$	$p > 0.448$	$\eta^2 = 0.006$
3. The workload in this unit was manageable	$F = 0.01$	$p > 0.945$	$\eta^2 = 0.000$
4. Assessment tasks in this unit were clear	$F = 0.34$	$p > 0.563$	$\eta^2 = 0.003$
5. The teaching staff gave me helpful feedback	$F = 3.96$	$p > 0.048$	$\eta^2 = 0.039$
6. The library resources met my needs for this unit.	$F = 0.12$	$p > 0.728$	$\eta^2 = 0.001$
7. I would recommend unit to other students	$F = 2.99$	$p > 0.086$	$\eta^2 = 0.030$
8. Technologies used performed satisfactorily	$F = 1.31$	$p > 0.254$	$\eta^2 = 0.013$
9. On-line enhanced my learning experience	$F = 2.56$	$p > 0.112$	$\eta^2 = 0.025$
10. This unit challenged me to learn.	**$F = 10.62$**	**$p < 0.002$**	**$\eta^2 = 0.098$**

Here a significantly lower mean rating from students studying units in online mode of offer (compared to students studying units offered conventionally) was observed only for SETU item 10 (**in bold**). The effect size here is modest – the observed difference in mean ratings based on mode of offer for SETU item 10 was 9.8 per cent. In the area of research into student engagement, a generally positive association between academic challenge and the use of information technology in student learning has been observed (Chen, Lambert, & Guidry, 2010; Nelson Laird & Kuh, 2005). However, in that context 'academic challenge' is conceived as a multi-dimensional benchmark scale that is composed of 11 items relating to levels of student participation in certain educationally purposeful activities, none of which actually mention the term 'challenge[d]' (Kuh, 2009). The simple form of SETU item 10 should not be seen as a direct measure of academic challenge in a student engagement context; in fact it is not really clear what it attempts to measure. Purposefully designed academic challenge and engagement in online learning environments has been identified as an important element in directing students to deep rather than shallow learning approaches (Garrison & Cleveland-Innes, 2005), and large-scale student engagement research has shown that institutions that focus on information technology mediated interactions and relationships may be associated with low levels of student academic challenge (Pike & Kuh, 2005). Though the effect size of this grouping is relatively small, this finding has significance for universities contemplating a heavy investment in the on-line modality, especially if the basis for the investment is to enhance students' overall learning experience and engagement.

Conclusions

It is important to acknowledge the limitations of this investigation. The available data permitted an investigation of class size, class level and online mode of offer status. Although class size and class level are commonly reported systematic influences on SET ratings, and online mode of offer has previously been found to be an additional systematic influence on SETU ratings at Deakin University, there are likely to also be other influencing factors not captured in this study. The available data did not permit a complete demarcation between the mean SETU ratings from on-campus, off-campus and wholly online enrolled students. However, the grouping employed to identify units with an online mode of offer (units with no on-campus enrolled students and units offered only in wholly online mode) is a realistic reflection of both the student experience and the practical use of SETU data during the period under consideration. While the initial groupings for class size were well balanced, the final two-way groupings used for class size, year level and online mode of offer in the

ANOVA tests were all significantly unbalanced. In practice, where only one grouping factor is being tested and the homogeneity of variance is accounted for (as was true in all the cases here), the ANOVA test is relatively robust to departures from equal group sizes. With only 100 sets of unit mean SETU ratings available to be included in the investigation, the data set is relatively small. However, even with a comparatively small data set, a number of differences in mean SETU ratings were observed at the $p < 0.02$ level of significance, lending weight to the conclusion that the observed differences are meaningful.

This paper presents an investigation into the systematic influence on SET ratings for engineering programs at Deakin University. Based on the mean SETU ratings for a sample set of 100 units of study, the following significant influences were observed. Class sizes with an enrolment of less than 35 reported significantly higher mean ratings for eight out of the ten SETU items – class size seemed to have a strong halo effect on most SETU items. Postgraduate level units reported significantly higher mean ratings for SETU items 7 – this item is important in as much as it is often taken as an overall proxy for SETU results. Units offered solely in online mode with no on-campus enrolment reported significantly lower means ratings for SETU item 10 relating to academic challenge. The observed effect sizes for significant differences in mean SETU item ratings were generally below 10 per cent. However, for SETU item 7, based on the school-wide mean rating of 3.46 out of 5 during the period under consideration, the observed effect size ($\eta^2 = 0.124$) translates to a possible difference in mean SETU rating of greater than 0.4 higher for a class size less than 35 compared to units with a larger enrolment, all other things being equal. This is a non-trivial impact in an environment where the first decimal place for unit SETU ratings might form part of the academic staff performance management system. More generally, while the use of SET data may have originally been primarily for formative purposes to improve teaching and learning, they are also increasingly used in many quarters for summative judgements of teaching quality and teaching staff performance that have implications for personnel decision making. 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. Fair and sound decision making based on SET data requires that the known systematic influences on SET data be accounted for.

References

Adams, Moya, Neumann, Ruth, & Rytmeister, Cathy. (1996). Is it a Level Playing Field? Factors Which Influence Student Evaluation of Teaching. Paper presented at the Australian Association for Research in Education Annual Conference, Singapore.

Allen, I. Elaine, & Seaman, Jeff. (2010). Class Differences Online Education in the United States, 2010. Babson Park, MA: Babson Survey Research Group / The Sloan Consortium.

Aron, Arthur, Aron, Elaine, & Coups, Elliot J. (2009). Statistics for Psychology (5th ed.). Upper Saddle River, N.J.: Pearson Prentice Hall.

Centra, John A., & Gaubatz, Noreen B. (2000). Is There Gender Bias in Student Evaluations of Teaching? The Journal of Higher Education, 71(1), 17-33.

Chen, Pu-Shih Daniel, Lambert, Amber D., & Guidry, Kevin R. (2010). Engaging online learners: The impact of Web-based learning technology on college student engagement. Computers & Education, 54(4), 1222-1232. doi: http://dx.doi.org/10.1016/j.compedu.2009.11.008

Davies, Martin, Hirschberg, Joe, Lye, Jenny, Johnston, Carol, & McDonald, Ian. (2007). Systematic Influences on Teaching Evaluations: The Case for Caution. Australian Economic Papers, 46(1), 18-38. doi: 10.1111/j.1467-8454.2007.00303.x

Deakin University. (2014). Unit Evaluation System. Retrieved 12 March, 2014, from http://www.deakin.edu.au/unit-eval/results/general.php

Denson, Nida, Loveday, Thomas, & Dalton, Helen. (2010). Student evaluation of courses: what predicts satisfaction? Higher Education Research & Development, 29(4), 339 - 356.

Garrison, D. Randy, & Cleveland-Innes, Martha. (2005). Facilitating Cognitive Presence in Online Learning: Interaction Is Not Enough. American Journal of Distance Education, 19(3), 133-148. doi: 10.1207/s15389286ajde1903_2

Hippensteel, Scott P., & Martin, Walter. (2005). Increasing the Significance of Course Evaluations in Large-Enrollment Geoscience Classes. Journal of Geoscience Education, 53(2), 158-165.

Jaccard, James, & Wan, Choi K. (1996). LISREL approaches to interaction effects in multiple regression. Thousand Oaks, Calif.; London: SAGE.

Johnson, Michael D., Narayanan, Arunachalam, & Sawaya, William J. (2013). Effects of Course and Instructor Characteristics on Student Evaluation of Teaching across a College of Engineering. Journal of Engineering Education, 102(2), 289-318. doi: 10.1002/jee.20013

Kember, David, Leung, Doris Y. P., & Kwan, K. P. (2002). Does the Use of Student Feedback Questionnaires Improve the Overall Quality of Teaching? Assessment & Evaluation in Higher Education, 27(5), 411 - 425.

Kuh, George D. (2009). The national survey of student engagement: Conceptual and empirical foundations. New Directions for Institutional Research, 2009(141), 5-20.

Loveland, Karen A. (2007). Student Evaluation of Teaching (SET) in Web-based Classes: Preliminary Findings and a Call for Further Research. The Journal of Educators Online, 4(2), 1-18.

Marsh, Herbert W., & Roche, Lawrence. (1993). The Use of Students' Evaluations and an Individually Structured Intervention to Enhance University Teaching Effectiveness. American Educational Research Journal, 30(1), 217-251.

Mayadas, A. Frank, Bourne, John, & Bacsich, Paul. (2009). Online Education Today. Science, 323(5910), 85-89. doi: 10.1126/science.1168874

Nelson Laird, Thomas, F., & Kuh, George D. (2005). Student Experiences with Information Technology and Their Relationship to Other Aspects of Student Engagement. Research in Higher Education, 46(2), 211-233. doi: 10.2307/40197353

Neumann, Ruth. (2000). Communicating Student Evaluation of Teaching Results: Rating Interpretation Guides (RIGs). Assessment & Evaluation in Higher Education, 25(2), 121-134.

Norman, Geoff. (2010). Likert scales, levels of measurement and the "laws" of statistics. Advances in Health Sciences Education, 15(5), 625-632. doi: 10.1007/s10459-010-9222-y

Palmer, Stuart, & Smith, Calvin. (2013). Updating RIGs: including the systematic influence of online study on student evaluation of teaching. Educational Research and Evaluation, 19(1), 77-90. doi: 10.1080/13803611.2012.746712

Pike, Gary R., & Kuh, George D. (2005). A Typology of Student Engagement for American Colleges and Universities. Research in Higher Education, 46(2), 185-209. doi: 10.2307/40197352

Ragan, James F., & Walia, Bhavneet. (2010). Differences in Student Evaluations of Principles and Other Economics Courses and the Allocation of Faculty across Courses. The Journal of Economic Education, 41(4), 335-352. doi: 10.1080/00220485.2010.510389

Rovai, Alfred P., Ponton, Michael K., Derrick, M. Gail, & Davis, John M. (2006). Student evaluation of teaching in the virtual and traditional classrooms: A comparative analysis. The Internet and Higher Education, 9(1), 23-35. doi: DOI: 10.1016/j.iheduc.2005.11.002

Santhanam, Elizabeth, & Hicks, Owen. (2002). Disciplinary, Gender and Course Year Influences on Student Perceptions of Teaching: Explorations and implications. Teaching in Higher Education, 7(1), 17-31. doi: 10.1080/13562510120100364

Smith, Calvin. (2008). Building effectiveness in teaching through targeted evaluation and response: connecting evaluation to teaching improvement in higher education. Assessment & Evaluation in Higher Education, 33(5), 517-533.

Copyright statement