

## **Online student evaluations of teaching: what are we sacrificing for the affordances of technology?**

In the context of increased emphasis on quality assurance of teaching, it is crucial that student evaluation of teaching (SET) methods are both reliable and workable in practice. Especially, online SET tends to raise criticisms with those most reactive to mechanisms of teaching accountability. However, most studies have been conducted with convenience, small and cross-sectional samples. Longitudinal studies are rare, as comparison studies on SET methodological approaches are generally pilot studies followed shortly after by implementation. The investigation presented here significantly contributes to the debate by examining the impact of the online administration method of SET on a very large longitudinal sample at course level. The study explores the impact of the administration method of student evaluations of teaching (paper based in-class versus off-class online collection) on scores with a longitudinal sample of over 63,000 student responses collected over a total period of ten years. Having adjusted for the confounding effect of class size, faculty, year of evaluation, years of teaching experience and student performance, it is observed that the actual effect of the administration method exists, but is insignificant.

Keywords: student evaluation of teaching, online survey, Ireland, response rates, educational technology

### **Introduction**

It is generally accepted that Universities are being held responsible for how well they serve the student population, and it has become common practice in universities for students to evaluate their lecturers. Student evaluation of teaching (SET), the most commonly used method of assessing a teacher's effectiveness, offers important opportunities for feedback and development but have their limitations, especially when they represent the only method of teaching evaluation. Used in conjunction with other methods of performance assessment, for example peer evaluations and teaching portfolios, student ratings may lead to improvement in teaching particularly if the teacher shares the results with a colleague or a teaching advocate. Over twenty years ago, Theall and Franklin (1990) identified student evaluations of teaching

as the single largest area of research in post-secondary education research; and less than ten years later, McKeachie and Kaplan (1996) identified over two thousand articles on this research topic. SET are by now a ubiquitous method for ‘measuring’ teaching effectiveness in many institutions, though there are a number of issues that have been highlighted in relation to how they are administered and their impact. Benson and Lewis (1994) review some of these negative attitudes and stereotypes in relation to SET initiatives. This summary is still relevant in relation to criticisms levelled at SET processes across the literature; at the more benign end of the scale, SET is characterised as a ‘necessary evil’(see also Ory 1991), at the more sinister, as a tangible illustration of administrative interference which ‘illustrates the mercantile philosophy of consumerism’ (Benson and Lewis 1994, p. 195) and a culture of bureaucratic surveillance (e.g. Johnson 2000). Many criticisms radiate from the contention that students are not a reliable source of information about teaching; for example, teachers may be rated highly because they are popular rather than on the basis of their content knowledge (e.g. Tomasco 1980); students’ ratings may be influenced by the grades that they receive: the higher the grades, the better the rating will be (Greenwald and Gilmore 1997; Neath 1996). At a more fundamental level, research literature focused early on the validity and reliability of the instruments used to elicit students’ reactions to teaching (Abrami 1989; Abrami *et al.* 1990; Cohen 1981). Potential sources of bias in terms of response to teachers from the point of view of gender, age or perceived physical attractiveness have also been investigated (Ambady and Rosenthal 1993; Hamermesh and Parker 2003; Kaschak 1978; Sinclair and Kunda 2000). In the case of Zabaleta (2007), a range of potential ancillary variables that could impact on SET, such as class size, gender, age, experience of teachers, are examined. Moore and Kuol (2005a) acknowledge controversies in the use of SETs, and the possibilities of bias from a variety of different quarters, but also make a strong case for using SET. The authors collect evidence suggesting that students can provide useful

information about their experience of teaching and learning that is preferable, at least, to ‘the proliferation of unrepresentative information and feedback about teaching relying on hearsay and anecdote’ (ibid.). Moreover, the greater benefit of using a student evaluation of teaching system is the explicit valorisation of the teaching component of the academic’s professional role and how this contributes to a move towards ‘parity of esteem’ (p. 60) with regard to research activities. In the context of this debate, one essential precondition for faculty to consider seriously is the feedback provided through a SET system that is perceived to provide reliable and valid information. Their trust is however challenged most often when SET relies on voluntary participation by students through an online survey system, as is the case in most universities currently.

### ***The reliability challenge of online student evaluations of teaching***

There are obvious reasons behind decisions to move existent evaluation systems online – the time savings in terms of administration and processing of feedback being the most compelling. Time efficiencies involve savings in terms of administration, but also for teachers and students who, naturally, have many and competing demands on their time. With online SET, not only their class time is preserved, but students also have longer to reflect on their answers, a factor which they tend to appreciate (Dommeyer *et al.* 2004; Layne *et al.* 1999). Provided that students have access to technology and are able to use it, the fact that students can reflect on their responses at relative leisure can impact positively the quantity and quality of written comments (Hardy 2003; Sorenson and Reiner 2003). Despite these benefits, where a changeover has occurred from long-established paper-and-pencil systems to online ones, this may engender a cultural change in the institution. Some students may be concerned that their anonymity may be compromised and their identity tracked by new online systems, while faculty may also wonder about confidentiality and who may have access to

and control of their evaluations (c.f. Sorenson and Reiner 2003). These conditions are central to the philosophy of the system which is to provide formative feedback for the purpose of facilitating pedagogical awareness and enhancements (Keane and Mac Labhrainn 2005; Ramsden 1991). But by far, the main reason why the progressive transition towards online administration methods in the last decade has challenged the credibility of SET relates to sample representation. It is widely acknowledged that, provided there is high class attendance –which in itself is a substantial assumption–, traditional paper-and-pencil administrations tend to yield higher response rates (Avery *et al.* 2006; Bennett and Nair 2010; Dresel and Tinsner 2008; Morrison 2011; Nulty 2008; Perrett 2011; Sax *et al.* 2003). This is due, of course, to the use of a ‘captive audience’ targeted during in-class evaluations, where the instructor or personnel from an educational development centre circulate and collect evaluation forms. These issues bring to light the importance of investing in educational research that informs practice and reassures sceptical faculty about the quality of the feedback that they receive.

Some existing research has explored some of the potential bias that online collection may have over SET scores. The general consensus is that, while the electronic administration method almost invariably results in lower response rates, its impact on results is non-existent or minimal (Ardalan *et al.* 2007; Avery *et al.* 2006; Dommeyer *et al.* 2004; Dresel and Tinsner 2008; Johnston 2003; Layne *et al.* 1999; Liegle and McDonald 2005; Perrett 2011; Stowell *et al.* 2011; Winer and Sehgal 2006). However, some recent conflicting evidence has also been presented (Morrison 2011; Nowell *et al.* 2010) that reports lower average SET ratings from evaluations conducted online. As a result, these authors argue that moving to online evaluation is not overly problematic, yet it is not comparable to in-class administration and as a result, institutions should choose one method or the other. Meinefeld (2010) goes beyond by deeming this administration method as convenient, but unreliable. These contradictory findings tend to find home with those most reactive to mechanisms of teaching

accountability, and in any case highlight the importance of being cautious about generalising findings from elsewhere into particular institutional contexts.

Despite the coverage of the issue in the literature, most studies have been conducted with convenience, small and cross-sectional samples and longitudinal studies are rare, as comparison studies between online and paper evaluations are generally pilot studies followed shortly after by implementation (Perrett 2011). Moreover, most studies are conducted at the student unit rather than course level, assuming that student responses can be considered independent, when their contextual circumstances –including having the same instructor–, actually makes them interdependent (Blair et al, 1983; in Perrett 2011). The investigation presented here significantly contributes to the debate by examining the impact of the online administration method of SET on a very large longitudinal sample at course level, while controlling for the known confounding effect of factors such as class size, faculty of study, year of evaluation, years of teaching experience and student performance.

### ***Background***

In the Irish university where this study took place (student registration around 13,000) teaching quality is considered to be of pivotal strategic importance in order to provide an outstanding and distinctive experience for every student. A majority of those staff who participated in the most recent quality survey (Quality Support Unit, University of Limerick 2011) claim to be concerned about teaching quality, welcome the requirement of the European Standards and Guidelines (ENQA 2009) to review it, and are also supportive of the idea that teaching quality should be systematically reviewed on a frequent basis. The Centre for Teaching and Learning (CTL) at the institution provides a structured approach to getting feedback from students about individual approaches to teaching, with a focus on continuous professional development rather than accountability. Available each semester, the SET

system has been running since the turn of the new century when it started on a small scale, and has grown enormously, reaching almost 500 evaluations each semester (which we refer to here as ‘round’). The process is voluntary and confidential at all times, and is designed to provide useful information to individual lecturers on their students’ experiences of the modules they teach. A structured, standard evaluation providing individual lecturers and tutors with quantitative ratings on a total of 18 teaching and module related items is used (see Methods section). After data collection and analysis, lecturers are provided with a personalised comprehensive report that includes mean and median scores and percentiles for each of the items of the scale. Academic staff can compare their own ratings with the average of assessed lecturers in similar class sizes for their own faculty. Importantly, and in addition to quantitative ratings, the evaluation also includes qualitative comments from individual students, which are consistently scanned for defamatory or insulting comments. These are also incorporated into the report, and can help to provide a more fine-grained picture of the issues, challenges and competencies associated with each teacher’s performance. This is complemented with an objective analysis of student evaluations by an educational developer within the CTL. The formative evaluation process acts as a supportive system helping individual teachers to create a current picture of the teaching and learning environment through the eyes of their students, with the aim of subsequently addressing issues that are identified. It can also serve to endorse and reinforce the approach of excellent teachers. Recipients are encouraged to interpret feedback carefully, recognising moderating factors such as class size, subjects taught, student seniority and previous performance. Most importantly, academic staff are encouraged to see the SET process as part of a larger suite of professional development initiatives, such as engaging in peer assisted reflection or documenting their professional development within a reflective portfolio format (S. Moore and Kuol 2005a; Sarah Moore and Kuol 2005b).

The SET service is a labour intensive process requiring time and input from a team involving a manager and senior administrator, an educational researcher and an educational developer who provides an objective analysis of student evaluations and provides further pedagogical guidance and development opportunities. Initially, the survey instrument was distributed and collected during class time with the help of independent support staff. All surveys were then processed manually through a reader scanner, which proved to be an extremely time intensive procedure. As demand for the service increased, and due to resource constraints, an online administration method through the students' email address was considered six years later from round 14. A successful pilot followed phased in implementation for a period of two academic years, during which lecturers were given the option to have an in-class or off-class (online) evaluation. From Autumn semester 2009/10 (coinciding with round 18) all evaluations were conducted online and the use of paper based in-class administration was eliminated. Since, many of our teachers have stated their appreciation for the efficiency of the new system as turnaround of reports is much faster and therefore, they are in a better position to make changes to their curricular design in the same semester. However, the lower response rates that are associated with online surveys have created a level of resistance by some members of staff. It has been anecdotally argued that self-selected responses can potentially distort SET scores, and that the sampling method is unfair as the online system offers every student equal access to the evaluation process regardless of class attendance. It is sometimes felt that absent students may be making unfair claims about teaching quality and that the online medium favours the expression of negative criticism to a greater extent than in-class data collection, as it is likely that the most engaged students are present on the day that the data is collected in the classroom. This complaint is justified by the fact that, despite its developmental focus, it has become frequent practice to request academics to include their SET reports in promotion procedures, and SET reports are

often requested for quality reviews within departments, faculties and across the institution (although the anonymity of personal teacher and module data is preserved at all times). It is for these reasons that it is important to explore to what extent the administration method (convenience in-class data collection versus self-selected online responses) may impact the SET scores in this particular context in order to inform the fairness of any potential comparisons.

## **Methodology**

### ***Sample***

This investigation is based on data collected during a period of ten years and includes a total of 88,686 observations. The most represented faculty is the Kemmy Business School, KBS (n= 37,620), followed by Science & Engineering, S&E (n= 24,941), Arts, Humanities & Social Sciences, AHSS (n= 14,944) and Educational & Professional Studies, E&PS (n= 10,973). About half this sample is comprised of students in large classes of registration over 100, about 22% of respondents represent medium sized classes of between 50 and 100, and about 26% of students came from small classes with registration of less than 50. About 36% of the sample is composed of first year students, second and third years constituted around 23% and 19% of the total respectively, followed with 22% of fourth year students. Data was collected on 673 individual members of staff, who had voluntarily applied for a student evaluation of one or more of their modules. In general, lecturers had no say in the administration method, except between rounds 14-17 (each round corresponding to an academic semester), when they were given a choice. As this could introduce a potential bias on results, all the responses in that time frame have been coded as 'missing' in the administration variable. This means that from the total sample of 88,686 respondents, we have comparable data for 63,173 observations: in-class (round 1-13 inclusive) includes



48,362 responses; and off-class (round 18-21) accumulated 14,811 observations. While previous studies at course level have been conducted with relatively small number of courses (Perrett 2011), this study includes data for a very large number of courses, totalling 1,028 courses with in-class administration, and 654 courses with off-class administration.

### ***Instrument***

The survey instrument was composed by an expert group of academics at the institution and redesigned after an initial pilot. The scale is constituted by three main dimensions, measured in a 5 point Likert scale from ‘strongly disagree’ to ‘strongly agree’. The first group of ten items refers to the effectiveness of the teaching; the next eight deal with issues related to the module organisation and logistical issues; and finally seven items address student performance:

#### Lecturer items

1. Overall, this lecturer is effective in teaching this module.
2. The lecturer is knowledgeable about the topics covered in this module.
3. The lecturer is well prepared for class.
4. The lecturer is interested in and enthusiastic about the subject.
5. The lecturer encourages me to find out more about the subjects in this module.
6. The lecturer communicates effectively in class.
7. The lecturer makes me feel free to ask questions, disagree or express my ideas.
8. The lecturer is good at explaining difficult material.

9. The lecturer evaluates my work in a way that helps me to perform better.

10. The lecturer is actively helpful when I have encountered difficulties.

#### Module Scale

1. The module objectives are clear.

2. The classroom facilities are appropriate.

3. The texts and course materials are valuable learning aids.

4. The course is organised and sequenced well.

5. The module has significantly increased my knowledge in the subject areas.

6. The module has significantly increased my skills in the subject areas.

7. The subject matter is relevant to my educational goals.

8. Overall, this is an effective module.

#### Student items

1. I felt sufficiently prepared for this module.

2. I find this academic subject quite easy.

3. I attended most or all of the required contact hours for this module.

4. I find this subject interesting.
5. I handed in all necessary coursework on time.
6. I did all I could to contribute to my learning and understanding of this module.
7. I have worked hard to succeed in this module.

Students are also requested to complete demographic information relating to age, gender, and year of study. Finally, students are strongly encouraged to elaborate on additional issues that have impacted their learning experience in the course. Each observation is completed with the information provided by the lecturer at the moment of application and which includes faculty; years in employment, years teaching the course evaluated, and class size. Both students and lecturers are assured of anonymity and confidentiality for the data provided, and they are informed that under the terms of the Ethics Research Committee at the institution, data will only be used elsewhere for strict research purposes.

### ***Research question***

Data analysis in this study is conducted attending to the scale average scores as these are used at the institution in the context of SET and quality reports, and are therefore at the core of much of the political debate on the issue. The null hypotheses are formulated as follows:

H0 (1): The administration method (in-class vs. off-class) does not meaningfully impact SET average results in the Lecturer scale.

H0 (2): The administration method (in-class vs. off-class) does not meaningfully impact SET average results in the Module scale.

In investigating the research question, we avoid a simplistic divide of student evaluations set by the use of technology per se, but propose a multivariate causal analysis that is contextualised in a particular institutional scenario. The variable under study, administration method (with 'in-class' coded as 1 and 'off-class' as 2), refers to the whole procedure by which SET has been managed in the institution and combines the impact of selection method, response rate and medium of delivery. The in-class paper administration samples selected by convenience usually had high response rates from students that attended class the day of the evaluation, missing out on the opinions of the absent students. The off-class collection implies another form of self-selected sampling as a consequence of its electronic delivery, and usually transfers into lower response rates.

In order to ascertain how much of the variable of average SET scores can be explained by the administration method, while controlling for a number of contextual variables, multiple linear regression analyses were carried out by forward stepwise method using the PASW Statistics package version 20. The contextual variables taken into account in the analysis include faculty (which constitutes four different variables for each of the four at the institution, this is, AHSS, EHS, KBS and S&E), class size, round in which evaluation took place, number of years teaching at university or at third level, number of years teaching the module for which a SET was requested, year of study at the time of the survey, and student attendance (addressed by the third item of the Student scale). Moreover, we follow Perrett's (2011) call for including perceived student performance in the course (as measured by items 1, 2, 4, 5 and 6 of the Student scale). As previously mentioned in the background to this study and as also noted by the author (Perrett 2011, p. 87), issues related to student engagement in the learning process are a potential source of contention regarding the reliability of online SET. The contribution of these variables to the predictive model is expected to shed much needed light on (a) the impact of students' academic motivation and

course performance on course and/or instructor ratings; and (b) the interaction between engagement in the learning process and the administration method of the teaching evaluation.

## Results

Preliminary exploration has shown that average scores reasonably adjust to normal distribution. Residual analysis has been found to be satisfactory and a reasonable fit of residuals was also confirmed. Significant linear regression models include a total of 13 variables, but administration method (in-class vs. off-class) is not included in any of the significant models of the stepwise analysis. Model 6 has been found to be the most parsimonious ( $R^2=0.303$ ,  $p<0.001$ ), as the rest of the variables in the subsequent models contribute less than 0.005 to the total  $R^2$ . Model 6, which explains about 30% of the variance of the Lecture scale scores, includes faculty of S&E and class size group (both with negative Beta coefficients); the items ‘I find this subject interesting’ and ‘I felt sufficiently prepared for this module’ of the student scale, faculty of EHS, and round in which evaluation took place (all four with positive Beta coefficients). Therefore, we can safely accept the null hypothesis (1) that the administration method (in-class vs. off-class) does not meaningfully impact SET average results in the Lecturer scale.

Table 1. Results of multiple regression analysis with Lecture average scores

*Table 1 about here*

Significant linear regression models that include 11 of the variables considered have been found for the Module Scale as dependent variable. The most parsimonious regression model found, in the sixth iteration of the stepwise analysis ( $R^2=0.300$ ,  $p<0.001$ ), includes three variables with negative Beta coefficients (faculty of S&E, class size group, and number of years teaching in university or at third level) and three which impact scores positively: ‘I find

this subject interesting', 'I felt sufficiently prepared for this module', and round in which evaluation took place. Administration method is included in model 8 (with  $R^2=0.305$ ,  $p<0.001$ ), but the addition of this variable to the model contributes only with a change of  $R^2=0.002$ . Moreover, having adjusted for the confounding effect of several confounding variables, the use of the off-class collection method accounts for a reduction of only .08 in average Module Scale scores (standardised beta coefficient=-0.088,  $p<0.001$ ). In conclusion, the impact of administration method in SET Module average scores is minimal and can be safely neglected in this case, therefore accepting the null hypothesis (2) that administration method (in-class vs. off-class) does not meaningfully impact SET average results in its Module scale.

Table 2. Results of multiple regression analysis with Module average scores

*Table 2 about here*

## **Discussion**

This analysis has explored the impact of the administration method on SET scores using very large number of courses in a concrete institutional context from a longitudinal perspective. Having adjusted for the confounding effect of class size, faculty, year of evaluation, years of teaching experience, year of study, student attendance and performance, it is observed that the actual effect of the administration method exists, but is extremely small. The compounding factors that constitute the administration method (selection method, response rate and delivery mode) have not been shown to have the decisive impact that is often assumed. The null hypotheses that the administration method (in-class vs. off-class) does not meaningfully impact SET average results in its Lecturer and Module scales are therefore accepted.

Despite the large sample size and the longitudinal nature of the data analysed, some inherent methodological limitations need to be acknowledged. Firstly, while other studies have taken an experimental approach that guarantees random sampling (see for example Stowell *et al.* 2011), the contextual variables interacting with the administration method were uncontrolled in our investigation. However, the multivariate regression analysis, the vast sample size, and the use of the course analysis rather than the analysis at the student level can go some way to compensate for this. Secondly, results should not be automatically generalised beyond this particular university context, especially as the use of different teaching evaluation instruments makes comparability virtually impossible. Thirdly, many other variables could greatly add to the predictive analysis, as much of the variance of scores is still unaccounted for, and it remains unknown to what extent the instructor variable – the very aim of any SET- actually can explain. Finally, as response levels were very varied, future research could be extended to include this aspect and study potential bias in the cases where minimum required response rates by class size (Nulty 2008) are not met. Lower response rates to online SET surveys can be seen as an issue of student engagement in their own learning experience, and indeed, may reflect the lack of class attendance that has become frequent across the board. The online delivery of evaluation methods can further add to this disengagement (Bennett and Nair 2010). The motivations that encourage online responses, how they relate to attendance patterns and if these students are more prone to provide more meaningful qualitative feedback can be further investigated in our context. Research elsewhere indicates that students may be unwilling to continue to provide feedback if they are not convinced that their feedback is listened to or acted upon (Bennett and Nair 2010; Leckey and Neill 2001), and they may need to be convinced that this is the case (Harvey 2003). Some existing literature points to ways in which response rates can be maximised (Bennett and Nair 2010; Nulty 2008; Sax *et al.* 2003). In order to engage students in the evaluation process,

Bennett and Nair (2010) suggest doing just this by outlining explicitly why feedback is elicited and showing in a concrete way how their feedback is incorporated at class, unit and programme level. It is important making student feedback ‘mean’ something through action and acknowledgment of the information students take the time to provide by closing the feedback loop; this is, explaining to students how their contribution makes a difference to their teaching.

## **Conclusion**

Issues of quality and accountability have surfaced continuously in discussion in the third level teaching environment of late, ever more so as budgeting constraints grow ever tighter and efficiencies-based discourses prevail. The notion of accountability in general, and teacher accountability in particular, are primarily located in discourses of quality assurance, university policy and state concerns. Solbrekke and Englund (2011) have noted the saliency of the semantics of the term (with its origin in accounting, along with that much-used term, ‘value added’), and its inherent idea of quantification of some kind, as being somehow at variance with academic discourse. There is, undeniably, an uneasy relationship between mechanisms of accountability often associated with external quality assurance, and academic autonomy (e.g. Findlow 2008). However, as Ramsden (2003, p. 211) pointed out a decade ago, ‘Evaluation for accountability has become an essential part of today’s university...[t]he days when students’ experiences and comparability of standards were in the background and unprofessional teaching behaviour was quietly tolerated have gone.’ This aligns with European guidelines in teaching quality assurance which states that institutions should have ways of satisfying themselves that staff involved with teaching of students are qualified and competent to do so, and that they should be available to those undertaking external reviews. These guidelines literally recommend that ‘institutions should provide poor teachers with



opportunities to improve their skills to an acceptable level and should have the means to remove them from their teaching duties if they continue to be demonstrably ineffective’ (ENQA 2009, p. 17). This emphasis on accountability has often lead to an increased interest in student evaluation of teaching that follows none of the recommendations for good practice, and can lead to the inevitable tension between the bureaucratic and responsive approaches to accountability (Gleeson and O'Donnabhain 2009). As Lynch nicely puts it (2007, in Gleeson 2009), ‘in education not everything that counts can be counted and not everything that can be counted counts’.

In the context of increased emphasis on quality assurance of teaching, it is crucial that research methods are both reliable and workable in practice. Diminishing resources should not be an excuse for biased research that may result in unfair comparisons. With this in mind, we have provided evidence that may inform practice and policy from a concrete institutional setting, and adds substantially to the existing body of literature. It is important to remember however that the SET process at this institution was created for, and continues to be, a formative exercise which was never intended as a quality assurance method, but as a tool within the context of a holistic, positive, and proactive approach to feedback on performance. This philosophy based on voluntary, confidential and formative participation sets it apart from merely bureaucratic quality assurance approaches and places the debate central to the opportunities that the system may offer for continuous professional development. In terms of mediating between the existence of external (or internal) accountability frameworks and the professionalisation of third-level teaching, we need to conceptualise and claim some middle ground. As educational researchers and developers, our stance to accountability is based on the value added for the teacher – in terms of professional development- and the students she or he serves.



## References

- Abrami, P. 1989. How should we use student evaluations to rate teaching? *Research in Higher Education* 30, no. 2: 221-27.
- Abrami, P., S. D'apollonia and P.A. Cohen. 1990. Validity of student ratings of instructions: What we know and what we do not. *Journal of Educational Psychology* 82, no. 2: 219-31.
- Ambady, N. and R. Rosenthal. 1993. Half a minute: Predicting teacher evaluations from thin slices of non-verbal behavior and physical attractiveness. *Journal of Personality and Social Psychology* 64, no. 3: 431-41.
- Ardalan, A., R. Ardalan, S. Coppage and W. Crouch. 2007. A comparison of student feedback obtained through paper-based and web-based surveys of faculty teaching. *British Journal of Educational Technology* 38, no. 6: 1085-101.
- Avery, R.J., W.K. Bryant, A. Mathios, H.J. Kang and D. Bell. 2006. Electronic course evaluations: Does an online delivery system influence student evaluations? *Journal of Economic Education* 37, no. 1: 21-37.
- Bennett, L. and C.S. Nair. 2010. A recipe for effective participation rates for web-based surveys. *Assessment & Evaluation in Higher Education* 35, no. 4: 357-65.
- Benson, D.E. and J.M. Lewis. 1994. Students' evaluation of teaching and accountability: Implications from the boyer and the asa reports. *Teaching Sociology* 22, no. 2: 195-99.
- Cohen, P.A. 1981. Student ratings of instruction and student achievement. *Review of Educational Research* 51, no. 3: 281-309.
- Dommeyer, C.J., P. Baum, R.W. Hanna and K.S. Chaplin. 2004. Gathering faculty teaching evaluations by in-class and online surveys: Their effects on response rates and evaluations. *Assessment and Evaluation in Higher Education* 29, no. 5: 611-23.
- Dresel, M. and K. Tinsner. 2008. Method effects of online gathered evaluations of teaching effectiveness in university courses. *Zeitschrift Fur Evaluation* 7, no. 2: 183-+.
- Enqa. 2009. Standards and guidelines for quality assurance in the european higher education area. Helsinki: European Association for Quality Assurance in Higher Education.
- Findlow, S. 2008. Accountability and innovation in higher education: A disabling tension? *Studies in Higher Education* 33: 313-29.
- Gleeson, J. and D. O'donnabhain. 2009. Strategic planning and accountability in irish education. *Irish Educational Studies* 28, no. 1.

- Greenwald, A. and G. Gilmore. 1997. Grading leniency is a removable contaminant of student ratings. *American Psychologist* 52, no. 11: 1209-17.
- Hamermesh, D. and A. Parker. 2003. Beauty in the classroom: Professor's pulchritude and putative pedagogical productivity, ed. 9853, NWPWP. Cambridge, MA: National Bureau of Economic Research.
- Hardy, N. 2003. Online evaluations: Fact and fiction. *New Directions for Teaching and Learning* 96: 31-38.
- Harvey, L. 2003. Student feedback. *Quality in Higher Education* 9, no. 1: 3-20.
- Johnson, R. 2000. The authority of the student evaluation questionnaire. *Teaching in Higher Education* 5, no. 4: 419-34.
- Johnston, T.D. 2003. Online student ratings: Will students respond? *New Directions for Teaching and Learning* 96, no. Winter: 49-59.
- Kaschak, E. 1978. Sex bias in the student evaluations of college professor. *Psychology of Women Quarterly* 2, no. 3: 235-42.
- Keane, E. and I. Mac Labhrainn. 2005. Obtaining student feedback on teaching & course quality: Briefing paper 2.
- Layne, B.H., J.R. Decristoforo and D. Mcginty. 1999. Electronic versus traditional student ratings of instruction. *Research in Higher Education* 40, no. 2: 221-32.
- Leckey, J. and N. Neill. 2001. Quantifying quality. The importance of student feedback. *Quality in Higher Education* 7, no. 1: 19-32.
- Liegle, J. and D. McDonald. 2005. Lessons learned from online vs. Paper-based computer information students' evaluation system. *Information Systems Education Journal* 3, no. 37: 3-14.
- Lynch, K. 2007. An egalitarian or a neoliberal concept of the educated citizen - examining the role of school league tables and university rankings in setting the agenda for education. *Journal of the National Association of Principals and Deputy Principals* 53-68.
- Mckeachie, W.J. and M. Kaplan. 1996. Persistent problems in evaluating college teaching. *AAHE Bulletin* 48, no. 6: 5-8.
- Meinefeld, W. 2010. Online-surveys for the evaluation of teaching at university-convenient and unreliable. *Kolner Zeitschrift Fur Soziologie Und Sozialpsychologie* 62, no. 2: 297-315.
- Moore, S. and N. Kuol. 2005a. A punitive bureaucratic tool or a valuable resource? Using student evaluations to enhance your teaching. In *Emerging issues in the practice of university learning and teaching*, eds O'neill, G, Moore, S and McMullin, B. Dublin: AISHE.

- Moore, S. and N. Kuol. 2005b. Students evaluating teachers: Exploring the importance of faculty reaction to feedback on teaching. *Teaching in Higher Education*, 10, Issue 1, 2005, Pages 57 - 73 10, no. 1: 57-73.
- Morrison, R. 2011. A comparison of online versus traditional student end-of-course critiques in resident courses. *Assessment & Evaluation in Higher Education* 36, no. 6: 627-41.
- Neath, I. 1996. How to improve your teaching evaluations without improving your teaching. *Psychological Reports* 78: 1363-72.
- Nowell, C., L.R. Gale and B. Handley. 2010. Assessing faculty performance using student evaluations of teaching in an uncontrolled setting. *Assessment & Evaluation in Higher Education* 35, no. 4: 463-75.
- Nulty, D.D. 2008. The adequacy of response rates to online and paper surveys: What can be done? *Assessment & Evaluation in Higher Education* 33, no. 3: 301-14.
- Ory, J.C. 1991. Changes in evaluating teaching in higher education. *Theory into Practice* 30, no. 1: 30-36.
- Perrett, J.J. 2011. Exploring graduate and undergraduate course evaluations administered on paper and online: A case study. *Assessment & Evaluation in Higher Education* 38, no. 1: 85-93.
- Qsu. 2011. Quality support unit, confidential report for the institutional reviews of irish universities. Limerick: University of Limerick.
- Ramsden, P. 1991. A performance indicator of teaching quality in higher education: The course experience questionnaire. *Studies in Higher Education* 16, no. 2: 129-50.
- Ramsden, P. 2003. *Learning to teaching in higher education*. Ed., Trans. Vol. `Volume| `of *Series Title*|. Edition ed|. Number of Volumes vols|. City|: Publisher|. (Orig. ub. Original Publication.)| URL
- Sax, L.J., S.K. Gilmartin and A.N. Bryant. 2003. Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education* 44, no. 4: 409-32.
- Sinclair, L. and Z. Kunda. 2000. Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin* 25, no. 11: 1329-42.
- Solbrekke, T.D. and T. Englund. 2011. Bringing professional responsibility back in. *Studies in Higher Education* First published on: 31 January 2011 (iFirst).
- Sorenson, D.L. and M. Reiner. 2003. Charting the uncharted seas of online student ratings of instruction. *New Directions for Teaching and Learning* 96: 1-24.
- Stowell, J.R., W.E. Addison and J.L. Smith. 2011. Comparison of online and classroom-based student evaluations of instruction. *Assessment & Evaluation in Higher Education* 37, no. 4: 465-73.

- Theall, M. and J.L. Franklin. 1990. Student ratings in the context of complex evaluation systems. In *Student ratings of instruction: Issues for improving practice. New directions for teaching and learning*, eds Theall, M and Franklin, J. San Francisco: Jossey-Bass.
- Tomasco, A. 1980. Student perceptions of instructional and personality characteristics of faculty: A canonical analysis. *Teaching of Psychology* 7: 79-82.
- Winer, L.R. and R. Sehgal. 2006. Online course evaluation analysis report. Montreal: McGill University.
- Zabaleta, F. 2007. The use and misuse of student evaluations of teaching. *Teaching in Higher Education* 12, no. 1: 55-76.