
The Move to a System of Flexible Delivery Mode (Online V Paper) Unit of Study Student Evaluations at Flinders University. Management Issues and the Study of Initial Changes in Survey Volume, Response Rate and Response Level

TALITHA BENNETT AND DAVID DE BELLIS

Flinders University, Australia

Accepted by the *Journal of Institutional Research* as Best Paper Winner at 2009 AAIR Forum

Abstract

Like the vast majority of Australian universities, Flinders University (Flinders) collects feedback from students on the quality of teaching and learning through unit of study (topic) or classroom evaluations. Prior to 2009, survey instruments at Flinders were delivered via paper mode and in person to students in the classroom. In a drive for an improvement to administrative efficiency, the adoption of a new system in 2009 created the option of two modes of delivery. Instruments can now be delivered via either paper mode in class, identical in most aspects to the previous system, or online mode, where students fill out the survey instrument, typically outside of class, using the World Wide Web. The choice of delivery mode is at the discretion of the academic. Students have no choice; they receive the survey either via paper or online. This article examines a number of aspects of this change in process at Flinders. Some general aspects of managing the implementation of the new process are discussed. Take-up rates of paper mode versus online mode show that the overwhelming majority of surveys were conducted online in 2009. A dramatic increase in the volume of surveying was noted. Response rates fell by approximately 20% overall and some evidence of a change in the level of response was apparent (cause unknown). Findings at the institution and faculty level are presented where applicable.

Key Words: online survey, evaluations, response rate, higher education

Student classroom surveys are used as a standard instrument in higher education to collect feedback from students in order to evaluate the effectiveness of teaching and learning within individual subjects. The feedback is used in support of the improvement of pedagogy or subject content, and typically plays a key role in the performance management of teaching staff.

This paper was first presented at the Annual Conference of the Australasian Association for Institutional Research, *Turning water into wine: Institutional Research as an essential ingredient in strategic decision making for tertiary education* held in Adelaide, 11–13 November 2009.

Correspondence to: David De Bellis, Head, Planning Services Unit, Flinders University, GPO Box 2100, Adelaide 5001, Australia. E-mail: david.debellis@flinders.edu.au

Although Nulry (2008) reports that surveying practice of this kind in Australia still varies, it is consistently reported in the literature that the prevalence of online delivery of student higher education classroom surveys throughout the world is ever increasing (Ardalan, A., Ardalan, R., Coppage, & Crouch, 2007; Ballantyne, 2003; Hoffman, 2003; Johnson, 2003; Lieberman, Bowers, & Moore, 2001; Mitra, Jain-Shukla, Robbins, Champion, & Durant, 2008; Nulry 2008; Sorenson & Reiner 2003). The ease of use of the web as an interface to collect and manage student responses has made the process of online surveying highly attractive.

The advantages of online surveying have been well documented (Ardalan et al., 2005; Layne, DeCristoforo, & McGinty, 1999; Leung & Kember, 2005; Mitra et al., 2008; Nulry, 2008; Sorenson & Reiner, 2003), including the:

- elimination of delays, costs and transcription errors associated with manual survey administration and data entry processes
- removal of the necessity to conduct surveys in class, leaving more time for teaching
- support for a real-time missing response check function if desirable
- efficiency and flexibility of improved reporting functions
- ability to deliver surveys to geographically remote students
- tighter control of the survey administration process eliminating the risk of indirectly or directly influencing results
- potential ability to link responses to external data sources to facilitate institutional research
- opportunity for all students to respond (not just those in class at the time of administering the paper survey)
- improvement in the quantity and quality of open-ended responses since students have proper time to consider their response
- potential for the reduction of impact on the environment by reducing the volume of paper.

However there have also been a number of disadvantages of online surveying quoted (Ardalan et al., 2005; Johnson, 2003; Layne et al., 1999; Leung & Kember, 2005; Mitra et al., 2008; Nulry, 2008) including:

- concerns from students around confidentiality and usability
- development and maintenance costs of an online software system
- facing the change in culture by changing a well-established practice
- response-level bias because, while the instruments are identical, the environments within which they are completed are quite different, potentially resulting in different respondent behaviour
- response-level bias introduced by the differing characteristics and attitudes of students between modes of delivery. This is particularly relevant in situations where response rates are low and the risk of respondents being non-representative is high.

There is, seemingly, no mention in the literature of the lack of ability of paper-based surveying systems to control and monitor compliance in terms of frequency and coverage of surveys throughout an institution. If the question of whether to survey or not to survey is left entirely to the individual teacher, there is a risk that only the better

teachers would tend to initiate a survey process. While this question is invariably left to the teacher regardless of delivery mode (as it currently is at Flinders), there is a greater potential for control and monitoring of compliance when using centralised systems delivering surveys online.

This article looks at the changes to Flinders' administrative processes, survey volume, response rates, and level of response in moving from a paper-based to an online system of classroom surveying in the first half of 2009. It compares the Flinders experience with what has been typically reported in the literature. An understanding of these differences and the investigation of any atypical behaviour is crucial given that student feedback is an integral part of Flinders' teaching and learning quality management processes.

Background

From 1991 to 2008, Flinders has conducted classroom evaluations using the same system of surveying via paper forms. Flinders staff were responsible for the initiation of surveys, printing and photocopying of survey forms, delivery/collection of forms from students, and checking of contextual information prior to the final delivery of forms to an external provider. The external provider was contracted to supply the forms, enter the results (via automatic scanning), store and secure all data, and report in a standard way back to key Flinders stakeholders, including the staff initiating the survey. Until 2008, collation and checking of the returned forms has been the responsibility of the Staff Development and Training Unit. This responsibility was transferred to the Planning Services Unit in 2008.

Part of the 2007 Learning and Teaching Performance Fund money that had been awarded to Flinders University was used to undertake a project in 2007 entitled 'Student Evaluation of Teaching and Topics (SETs): Moving to a new system?'. This SET project examined whether content, mode and provider aspects of the Flinders system continued to meet the university's needs. The university endorsed one of the recommendations of the report, 'That the university phase in the introduction of an online evaluation system, to run parallel with the paper-based system'. The Planning Services Unit, governed by an advisory group, carried the implementation of a new web-based SET system under an outsourcing arrangement with the University of New South Wales, with a rollout date of January 1, 2009. The contract with the previous external provider was discontinued.

The SET system provides for both online and paper-based evaluation of teachers and topics. Selection of the mode is at the discretion of staff. Students have no choice in that they receive the survey either via paper or online. Online evaluations are completely and automatically processed through the system. Surveys are held open for a period designated by users. Students are automatically notified of the survey by email and subsequent periodic email reminders are sent. Once the survey is closed, results are immediate. Paper-based surveys are administered in much the same way as previously and this is still essentially a manual process. Flinders staff set up surveys in the system, print and photocopy forms, distribute and collect forms manually, and a third party supplies scanning and raw data delivery services. These raw results are then electronically imported and, from this point, reporting

mechanisms are identical to those for online surveys. The instruments (forms and questions) have not been reviewed and remain identical in structure and wording.

Administration Model

A clear plan for administering the new SET system needed to be developed well before the launch date to allow for proper planning and to ensure efficiency in process. After careful consideration of a number of alternatives, the following administrative framework was adopted.

Automatic Survey Pre-registration

Due to the ability of the new SET system to store all contextual survey registration information for individual classroom evaluations it became possible to implement a central automatic pre-registration of surveys. The pre-registration process alleviates most, if not all, of the data entry burden on users who would otherwise need to initiate and set up each individual survey from scratch within the system. A pre-registration specification for standard first-semester topics was negotiated individually with each Academic Organisation Unit (AOU) through, typically, the head academic. The specification included, in logical, structured, and systematically reproducible form, information on which topics were to be included (and excluded) for pre-registration, mode of delivery (online or paper), and survey open/close dates. In part, the specification formed an official pre-registration service agreement with the AOU. In total 2,596 surveys were centrally pre-registered as either a standard topic evaluation or standard teacher evaluation during the first half of 2009. The relevant topic coordinators were notified by email of the pre-registration and reminded to review survey parameters well before the open date.

Local and Central Support Structure

Nominated administrative staff, set up as power users with particular access rights within the system, are responsible for specific local support functions to academics within AOU's or other structural areas. Similar to the way surveys used to be administered, academics can still be responsible for managing their own surveys, operating the system themselves, and administering the paper form process if applicable. Decisions on the arrangements within each area with respect to how the responsibility for maintaining the classroom evaluations process is managed is left up to each individual area. Nonetheless, each AOU does have at least one (and typically more) staff who are responsible for local support. The Planning Services Unit provides a centralised support function to the whole university, but the burden on this unit is lessened by the local delegation of most of the support and administration responsibility.

Training

The extra power introduced by the adoption of the new SET system has inevitably increased the need for formal training. For basic administration of surveys it would be expected that the system is intuitive enough for a computer literate user to operate it successfully without formal workshop-type training. However, training services have been offered centrally by the Planning Services Unit. In the first half of

2009 the Planning Services Unit has run 14 half-day computer-based training workshops (79 staff attended), seven one-hour open-door lecture demonstrations (video also made available on the web), plus a small number of special training sessions for particular areas. One-on-one report retrieval skill sessions have also been offered to all heads of AOU.

Volume of Surveying

With the move to the new system of classroom evaluations in 2009, both the number of teacher and topic evaluations have more than doubled. The number of topic evaluations has increased from 450 in the first half of 2008 to 1,150 in the first half of 2009 (156% growth). The number of teacher evaluations has increased from 868 in the first half of 2008 to 1,780 in the first half of 2009 (105% growth).

Figure 1 shows the change in volume of topic evaluations from 2008 to 2009 by faculty; similarly, Figure 2 for teacher evaluations. While all faculties have increased their volume of surveying for both types of evaluations, Health Sciences (HS) and Science and Engineering (SE) (both faculties of the ‘hard’ sciences) have shown the greatest increase.

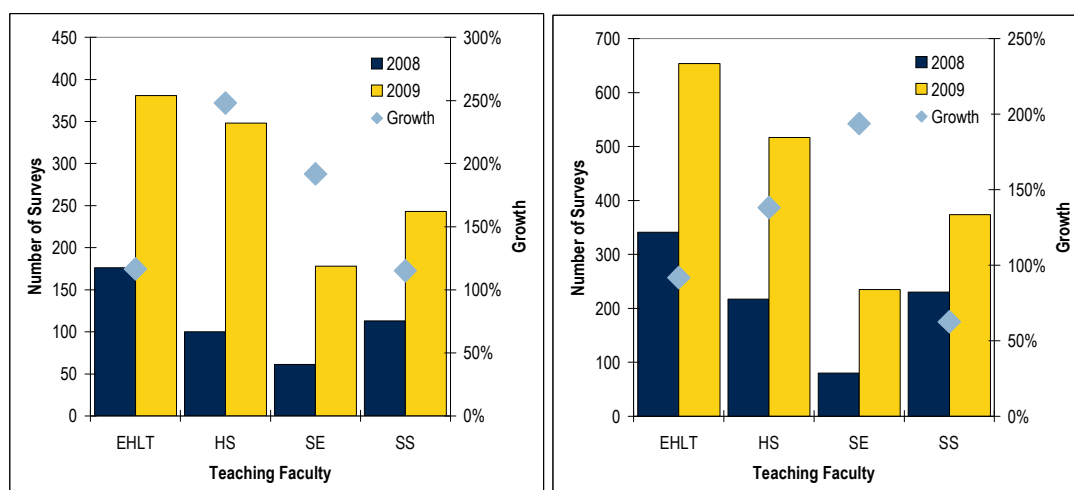


Figure 1: Number of topic evaluations by teaching faculty first-half 2008–2009.

Figure 2: Number of teacher evaluations by teaching faculty first-half 2008–2009.

Legend: EHLT = Education, Humanities, Law & Theology, HS = Health Sciences, SE = Science & Engineering, SS = Social Sciences.

It would be logical to conclude that the increase in volume of surveying is largely attributable to the introduction of a system of central automatic pre-registration of surveys. The vast majority of pre-registered surveys were set up to be conducted online, requiring no further effort from staff. It is impossible to know whether the general level of volume of surveying would have been affected by the move to a system of flexible mode delivery in the absence of automatic pre-registration. Nothing in the literature has been found on this subject.

Survey Mode Take-Up Rates

The vast majority of first half-year 2009 surveys (88%) were conducted online. Figure 3 shows the take-up rate of online and paper mode surveys for the first half of 2009 by faculty. The relatively high proportion of paper mode surveying in Education, Humanities, Law and Theology (EHLT) was at least partly due to non-standard teaching structures and students with special needs in language and theology topics where online evaluation is problematic.

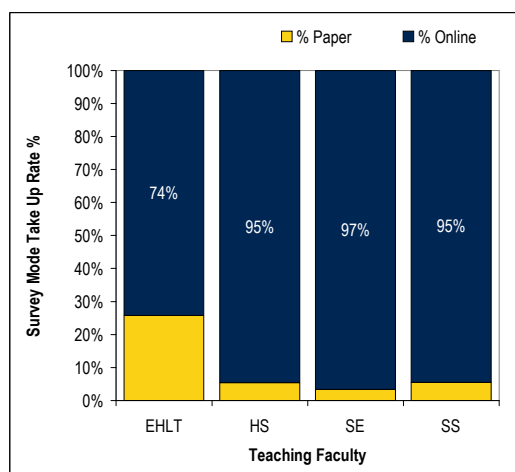


Figure 3: Take-Up Rate of Online and Paper Evaluations by Teaching Faculty First-Half 2009. EHLT=Education, Humanities, Law & Theology, HS=Health Sciences, SE=Science & Engineering, SS=Social Sciences.

Response Rates

As a new user of online mode of delivery of classroom surveys in 2009, response rates and the associated unmeasurable non-response bias are an obvious concern. Response rates from online surveys are influenced by a range of factors, including access to computers and the internet, computer literacy, and culture (Leung & Kember, 2005).

The literature suggests that out-of-class online evaluations of classroom teaching within tertiary institutions generally achieve response rates that are lower than in-class paper-based evaluations. Ardalan et al. (2007) reported an online response rate of 30.7% versus 68.7% for paper-based delivery in their study. Similarly, in an experiment looking at the effect of delivery mode on response ratings, Layne et al. (1999) reported an online response rate of 47.8% and a paper-based response rate of 60.6%. More recently, Nulry (2008) presented a summary of the literature, looking at publications from 1999 to 2006, and reported that all but one found a lower online response rate. Overall, the difference in response rates from these 8 studies was 23% (56% paper-based versus 33% online).

Table 1: Response rates of online and paper-based surveys at Flinders by teaching faculty first half 2008, 2009

	Faculty	Overall Response Rates			
		Online 2009	Paper-based 2009	All modes 2009	Paper-based 2008
Topic Evaluations	Education, Humanities, Law, & Theology	31%	52%	39%	52%
	Health Sciences	34%	53%	35%	62%
	Science & Engineering	31%	88%	31%	50%
	Social Sciences	34%	83%	37%	50%
	Flinders	33%	53%	36%	54%
Whole of Class Teacher Evaluations	Education, Humanities, Law, & Theology	35%	47%	44%	55%
	Health Sciences	29%	33%	29%	52%
	Science & Engineering	32%	72%	32%	46%
	Social Sciences	33%	48%	35%	49%
	Flinders	32%	44%	35%	51%

The overall response rates by teaching faculty for all online topic and whole-of-class teacher evaluation surveys conducted during the first half of 2008 and 2009 at Flinders are summarised in Table 1. Flinders experienced a very typical drop in the overall response rate of approximately 20% for online 2009 surveys compared with paper-based 2008 surveys. A degree of variation in response rates between faculty and type of survey (topic versus teacher) is noted. Response rates for 2009 paper-based surveys varied considerably between faculties due to the low take-up rate of the paper option and, therefore, the relatively small numbers of students who had the opportunity to respond to surveys in this way.

Level of Response

With the adoption of a system of online delivery from a paper-based mode of delivery, it is crucial for institutions to understand whether the different method of administration is likely, in itself, to have a significant effect on survey results. Changes in results can be measured in a number of ways, including the level of response of quantitative ratings, proportion of respondents providing comments, the level of response of qualitative ratings (positive, negative or mixed; constructive or destructive), and the length of comments (Ardalan et al., 2007). The discussion here is limited to changes in the level of response of quantitative ratings.

Previous research from controlled experiments generally suggests that no significant difference in the level of response of quantitative ratings should be expected between delivery modes for large samples.

Leung and Kember (2005), in their study of a sample of 2,786 undergraduate students from a university in Hong Kong, concluded that changing to an online collection mode would not affect the comparability of ratings on a 5-point Likert scale. In a study by Ardalan et al. (2007) of 1,364 business and public administration student respondents, there was no evidence of a significant difference in mean response for any of the eight quantitative questions analysed. It was also reported that these findings were consistent with a large sample study conducted by Liegle and McDonald (2005).

Layne et al. (1999) analysed the responses of 2,453 students enrolled in undergraduate- and postgraduate- level courses at a south-eastern university in the United States. They concluded, based on 5-point Likert scale responses from 14 core questions, that the survey method (online versus paper) main effect was not statistically significant. However, if the sample size is small, it could reasonably be expected that there is a high risk of results not being representative (Nulry, 2008). This is the case regardless of mode of delivery but, as discussed previously, an online mode of delivery is likely to produce a much lower response rate and therefore potentially a greater bias. Nulry (2008) looks at this in more detail but a proper discussion is outside of the scope of this article.

Regardless of the sample size, there is no way of determining whether the introduction of an online mode of delivery in 2009 at Flinders has caused a real change in the level of response from paper-based surveys in 2008. No formal experimental design was in place. The data gathered was purely observational without any controls. So any changes to response could be due to the effect of any number of confounding factors. Indeed, it is possible that any observed changes could be caused by shifts in topic or teaching quality from 2008 to 2009 — changes in the very attribute the instruments are designed to measure. However, it can at least be established whether there were any significant changes or not regardless of cause. The following analysis is presented looking for changes in the level of response (on a 7-point Likert scale) to the two core questions ‘Overall I was satisfied with the quality of this topic’ (from topic evaluations) and ‘Overall, this person is effective as a university teacher’ (from teacher evaluations). Likelihood Ratio Chi-square (χ^2 LR) tests of independence were carried out to find evidence against the null hypothesis that the distribution of response is identical for 2008 and 2009 (significance level 5%).

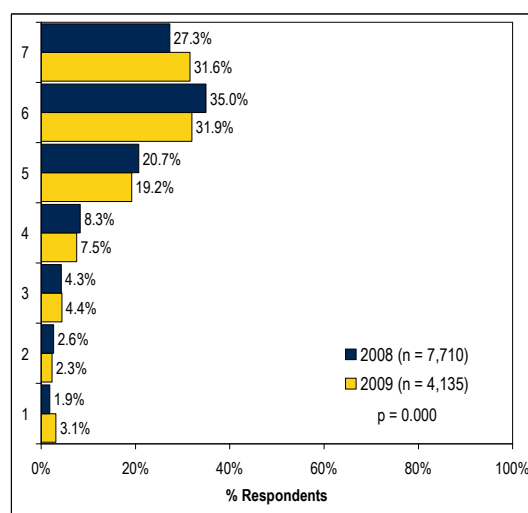


Figure 4: Distribution of responses on 7-point Likert scale to ‘Overall I was satisfied with the quality of this topic’ for 156 topics taught in both years at Flinders 2008 (paper mode) and 2009 (online mode).

1 = *Strongly Disagree*, 4 = *Undecided*, 7 = *Strongly Agree*.

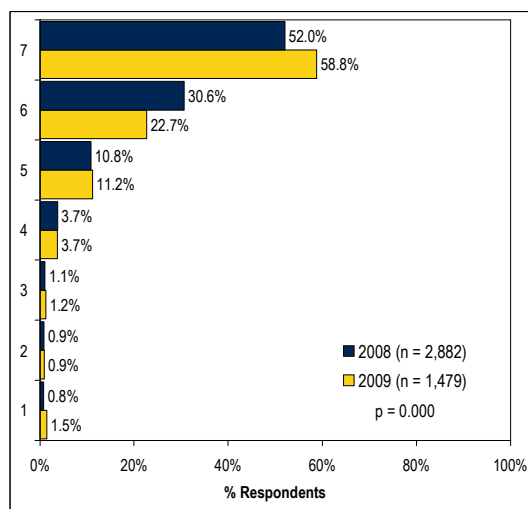


Figure 5: Distribution of responses on 7-point Likert scale to ‘Overall, this person is effective as a university teacher’ for 57 teachers who taught the same whole-class topics at Flinders 2008 (paper mode) and 2009 (online mode).

1 = *Strongly Disagree*, 4 = *Undecided*, 7 = *Strongly Agree*.

Figure 4 shows the difference in the distribution of response between 2008 (paper mode) and 2009 (online mode) on a 7-point Likert scale to the topic evaluation question ‘Overall I was satisfied with the quality of this topic’. Only topics (156) surveyed in both years were included to ensure comparisons were as meaningful as possible. Overall response rates were 52% and 32% for 2008 and 2009 respectively and typical of the rate profiles discussed in the earlier section. There is a statistically significant difference ($p = .000$) in the distribution of responses overall for 2009 compared with 2008 for Flinders overall. In 2009, students were more likely to respond with 1 (*Strongly Disagree*) or 7 (*Strongly Agree*), and less likely to respond with 6. The differences in the proportion of respondents on all other points of the 7-point scale were small or negligible. So while, in 2009, the likelihood of an extreme response has seemingly increased at both ends of the scale, the greater change is observed at the higher end. Therefore, there is evidence of an improvement in SET scores for overall topic satisfaction for Flinders in 2009 (online mode) compared with 2008 (paper mode) based on evaluations of these 156 topics.

Figure 5 shows the difference in the distribution of response between 2008 (paper mode) and 2009 (online mode) on a 7-point Likert scale to the teacher evaluation question ‘Overall, this person is effective as a university teacher’. Only whole class evaluations of the same teachers teaching the same topics (57) in both years were included to ensure comparisons were as meaningful as possible. Overall response rates were 47% and 33% for 2008 and 2009 respectively and again typical of the rate profiles discussed in the earlier section. There is a statistically significant difference ($p = .000$) in the distribution of responses overall for 2009 compared with 2008. Again, students were more likely to respond with 1 (*Strongly Disagree*) or 7 (*Strongly Agree*), and less likely to respond with 6 in 2009 compared with 2008. The differences in the proportion of respondents within all other points of the 7-point scale were small to negligible. So while the likelihood of an extreme response has seemingly increased at both ends of the scale for 2009, the greater change is observed

at the higher end. So there is evidence of an improvement in SET scores for overall teacher satisfaction for Flinders in 2009 (online mode) compared with 2008 (paper mode) based on results from evaluations of these 57 teachers.

The same tests on the same topic and teacher evaluations were repeated within each teaching faculty. Figure 6 shows the difference in the distribution of response between 2008 (paper mode) and 2009 (online mode) to the topic evaluation question ‘Overall I was satisfied with the quality of this topic’ for individual faculties.

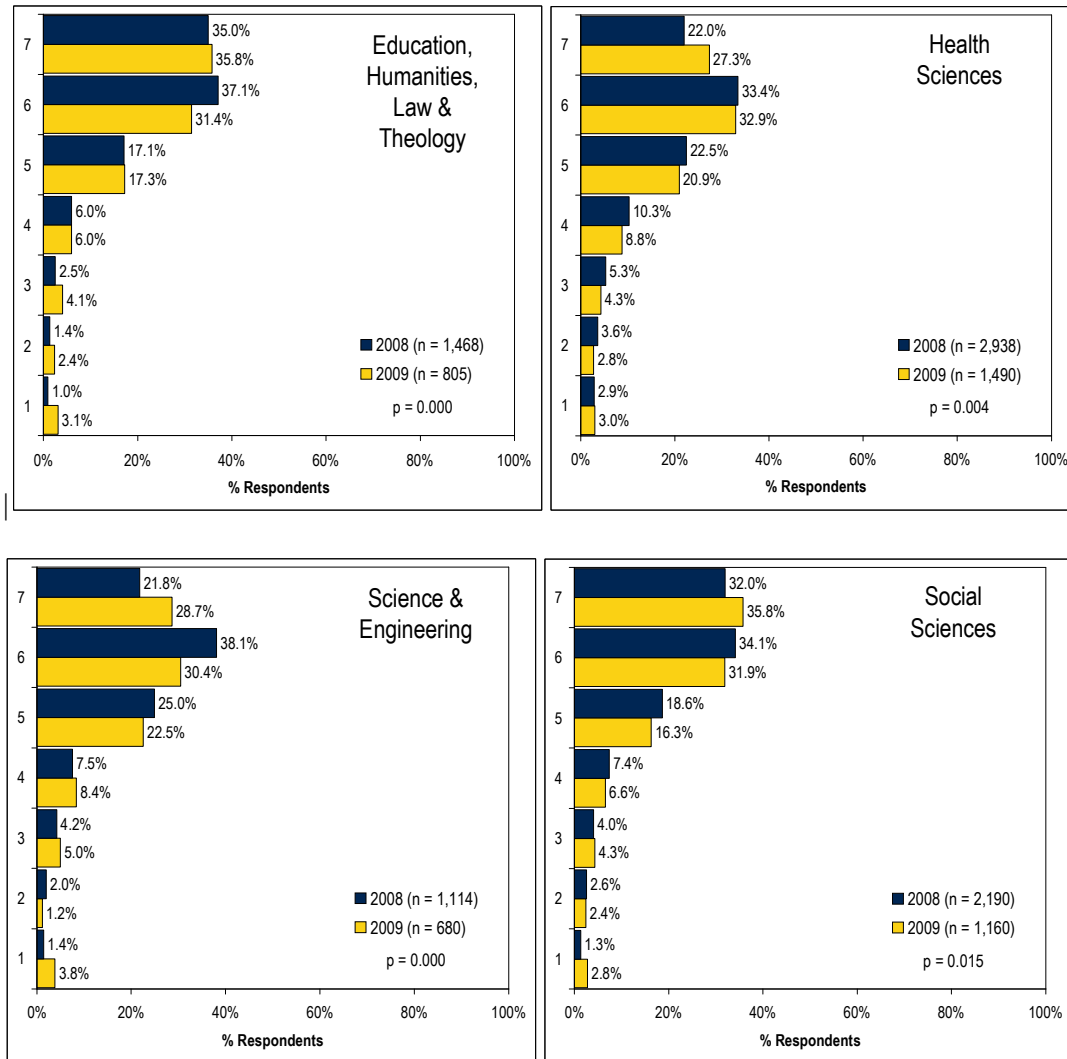


Figure 6: Distribution of responses on 7-point Likert scale to ‘Overall I was satisfied with the quality of this topic’ for 156 topics taught in both years at Flinders 2008 (paper mode) and 2009 (online mode) by teaching faculty.

1 = *Strongly Disagree*, 4 = *Undecided*, 7 = *Strongly Agree*.

While significant differences are noted in the distribution of response to topic satisfaction for all four faculties, the nature of these differences varies. Respondents in all faculties (except Health Sciences) were more likely to respond with 1 (*Strongly Disagree*). But, in general, there is evidence of an improvement in SET scores for

overall topic satisfaction within all faculties except Education, Humanities and Theology where some degradation is noted.

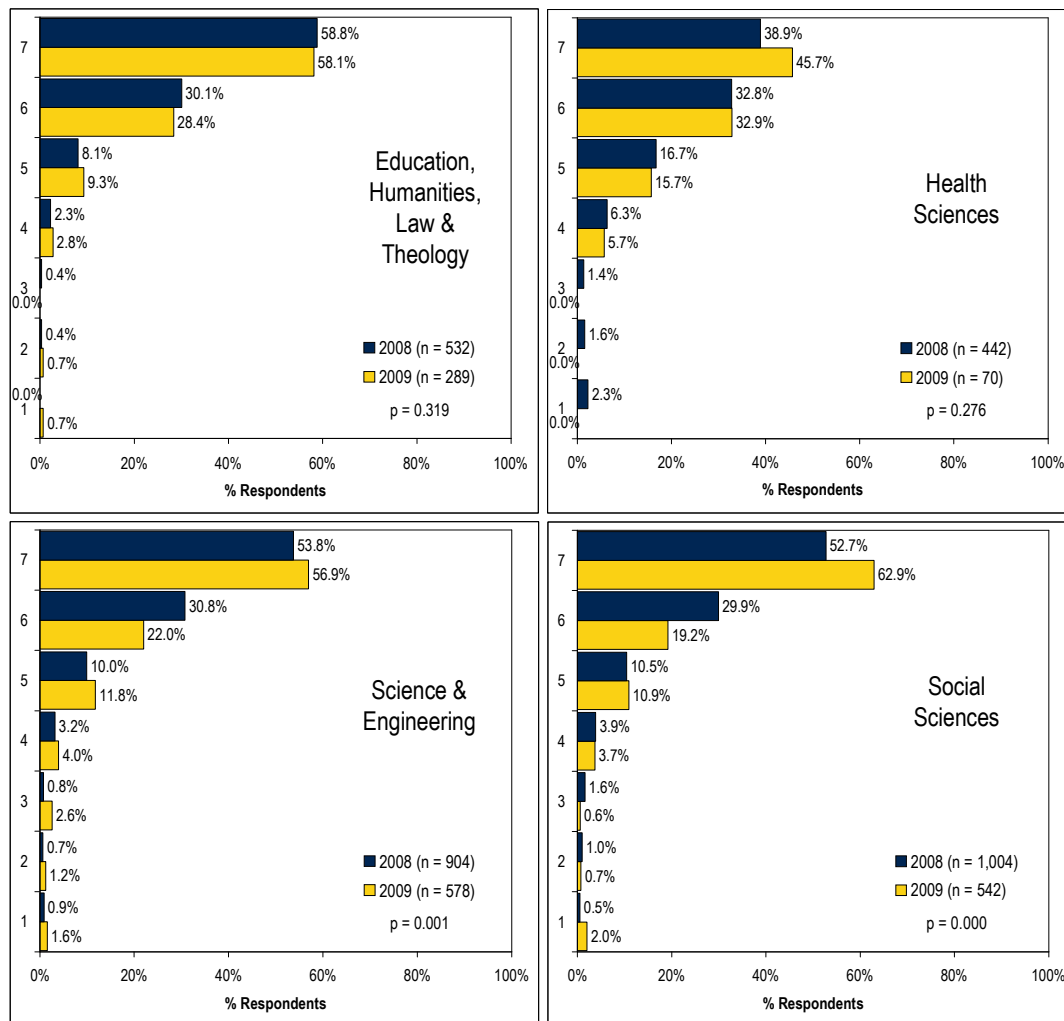


Figure 7: Distribution of responses on 7-point Likert scale to ‘Overall, this person is effective as a university teacher’ for 57 teachers who taught the same topics in both years at Flinders 2008 (paper mode) and 2009 (online mode) by teaching faculty.

1 = *Strongly Disagree*, 4 = *Undecided*, 7 = *Strongly Agree*.

Figure 7 shows the difference in the distribution of response between 2008 (paper mode) and 2009 (online mode) to the teacher evaluation question ‘Overall, this person is effective as a university teacher’ for individual faculties. There is no significant difference in the distribution of response for Education, Humanities, Law and Theology or Health Sciences. The interpretation for the remaining faculties, Science and Engineering and Social Sciences, is similar and influences/mirrors the whole of institution interpretation provided earlier.

Summary and Conclusions

A number of changes in administrative processes were required to support the transition from a purely paper-based model of delivering classroom evaluations in 2008 to a system of flexible delivery mode from 2009. Mechanisms for support and training, and the implementation of an automatic survey pre-registration system were key aspects of a management strategy that successfully supported users in 2009. The introduction of automatic pre-registration of surveys, which ensures little or no intervention is necessary by teaching staff to initiate and complete online surveys, has seemingly contributed to a marked increase in the number of surveys completed (the number of surveys have more than doubled). Therefore, while overall response rates have declined by approximately 20% (that is, 50% to 30% and quite typical based on the literature), the increase in survey volume has compensated. So overall, the total number of student responses received during the first half of 2009 is not dissimilar to the total number received during the first half of 2008 (of the order of 25,000).

Without proper experimental design it is not possible to determine if mode alone is responsible for any differences in student scores at Flinders in 2009 compared with 2008. The literature suggests differences of this kind would be highly unlikely. However, the distribution of student ratings overall were compared for a set of identical topics/teachers on the two main questions from the topic and teacher evaluation surveys. At an institutional level the conclusion for both questions was the same. That is, in 2009 the likelihood of an extreme response has seemingly increased at both ends of the ratings scale, but the greater change is observed at the higher end. Therefore, there is evidence of an improvement in SET scores for both overall topic and teacher satisfaction for the whole of Flinders during the first half of 2009 (online mode) compared with the first half of 2008 (paper mode).

Exact conclusions differ somewhat when ratings within each faculty are considered. However, except for Education, Humanities, Law and Theology topic ratings, there is either no evidence of any change in the distribution of scores, or there is some evidence that scores have improved. There is therefore little evidence, at least at the faculty level for the two core questions, of any wholesale degradation of SET scores for surveys conducted online in 2009 compared with those delivered via paper in 2008.

References

- Ardalan, A., Ardalan, R., Coppage, S., & Crouch, W. (2007). A comparison of student feedback obtained through paper-based and web-based surveys of faculty teaching. *British Journal of Educational Technology*, 38(6), 1085–1101.
- Ballantyne, C. (2003). Online evaluations of teaching: an examination of current practice and considerations for the future. *New Directions for Teaching and Learning*, 96, 103–112.
- Hoffman, K.M. (2003). Online course evaluation and reporting in higher education. *New directions for teaching and learning*, 96, 25–29.
- Johnson, T. D. (2003). Online student ratings: will students respond? *New Directions for Teaching and Learning*, 96, 49–59.

-
- Layne, B.H., DeCristoforo, J.R., & McGinty, D. (1999). Electronic versus traditional student ratings of instruction. *Research in Higher Education*, 40(2), 221–232.
- Leung, D.Y.P., & Kember, D. (2005). Comparability of data gathered from evaluation questionnaires on paper and through the internet. *Research in Higher Education*, 46(5), 571–591.
- Lieberman, D., Bowers, N., & Moore, D.R. (2001). Use of electronic tools to enhance student evaluation feedback. *New Directions for Teaching and Learning*, 87, 45–54.
- Liegle, J., & McDonald, D. (2005). Lessons learned from online vs. paper-based computer information students' evaluation system. *Information Systems Education Journal*, 3(37), 3–14.
- Mitra, A., Jain-Shukla, P., Robbins, A., Champion, H., & Durant, R. (2008). Differences in rate of response to web-based surveys among college students. *International Journal on Elearning*, 7(2), 265–281.
- Nulry, D.D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment & Evaluation in Higher Education*, 33(3), 301–314.
- Sorenson, D.L., & Reiner, C. (2003). Charting the uncharted seas of online student ratings of instruction. *New Directions for Teaching and Learning*, 96, 1–24.