

Does the timing of evaluations matter? An investigation into online student feedback and whether timing has an impact

Julie-Ann Pegden, Curtin University, Perth, Australia

Beatrice Tucker, Curtin University, Perth, Australia

Abstract

Anecdotal evidence suggests that some academic staff believe that student feedback varies depending on when students give their feedback. This study examined student feedback gathered in four different six week evaluation periods (Semesters 1 and 2 in 2009 and 2010) at Curtin University. Response rates and survey responses were analysed to determine when student subgroups give their feedback and whether student feedback differs according to week of survey submission. Results showed there was little variation in weekly response rates over the six week evaluation period according to student age, gender or semester weighted average, although there were lower rates of participation in week six by students of low semester weighted average than students of higher semester weighted average. Survey results varied slightly over the weeks and generally were lowest in weeks one and five and highest in weeks three and four. Differences in student feedback over the evaluation period were generally small and whilst there were some recurring patterns of variation in responses, there were also quite varied patterns of responses over different semesters in different weeks, particularly at the faculty level. Overall, week of survey submission did not have a consistent impact on survey results which were generally stable over time.

Key words: student evaluation of teaching and learning; student feedback; timing

This article has been peer-reviewed and accepted for publication in *SLEID*, an international journal of scholarship and research that supports emerging scholars and the development of evidence-based practice in education.

© Copyright of articles is retained by authors. As an open access journal, articles are free to use, with proper attribution, in educational and other non-commercial settings.

ISSN 1832-2050

Introduction

Student evaluations of subjects and teaching are increasingly important in Australian tertiary institutions. With the recent establishment of the Tertiary Education Quality and Standards Agency (TEQSA) it is anticipated that the results of student satisfaction surveys will become increasingly transparent with the publication of selected survey results on the My University Website (Gillard 2010).

The acceptance of student evaluations by academic staff is varied: many welcome student feedback and others remain guarded and mistrustful of student feedback (Oliver et al. 2001). Some academics feel that only the poorer performing students

give feedback using online evaluation surveys. While there is much speculation, relatively few student factors have been shown to bias student evaluations of teaching and learning and a review of all the evidence indicates that there is only a weak correlation between student evaluations of teaching and class-average grades (Marsh 2007).

Anecdotal evidence suggests that some academic staff at Curtin University believe that student feedback varies depending on when during the evaluation period students give their feedback. During professional development sessions with teaching staff academics often express their belief that students are not in a position to give feedback until they have completed all classes and assessments while the view of others is that that students who give feedback after sitting their final exam are influenced by the perceived level of difficulty of that exam. Some academics report that they actively discourage students from giving feedback until they have completed all classes and assessments.

Whilst it is generally accepted that student ratings are stable over time (Marsh 1987) and that ratings are highly reproducible for the same academic over time (Marsh and Hocevar 1991, Krantz-Girod et al. 2004), limited research has been conducted on the impact of timing on student evaluations and the results of these studies vary. Leigle and McDonald (2005) found that student feedback was more positive after exams and after students' grades had been released. They speculated that this was because students were more stressed before exams and this stress had a negative impact on evaluations. They also speculated that student evaluations were more positive after students knew their grades because they no longer had a 'fear of the unknown'. Results of this study showed that academically weaker students were less likely to give online feedback after their final exams. Coffey and Gibbs (2000) found student feedback was more positive at the end of semester than it was early in the semester, regardless of when students sat their final exams.

Method

In 2006, Curtin University implemented a university-wide student evaluation system called *eVALUate*. The *eVALUate* unit survey contains eleven quantitative items which ask students for their perceptions of what helped their achievement of unit learning outcomes (Items 1 to 7), their engagement and motivation (Items 8 to 10) and overall satisfaction (Item 11) (see Appendix for the *eVALUate* unit survey items). Students indicate Strongly Agree, Agree, Disagree, Strongly Disagree or Unable to Judge for each item. The survey items and response scale have been rigorously tested and found to be reliable and valid (having content validity and face validity with Australian and International students) (Oliver et al. 2008). Statistical testing shows the rating scale is sufficiently discriminating to identify areas of teaching and learning practice that need attention (Oliver et al. 2008).

The *eVALUate* unit survey is administered online through the student web portal. Students are notified when *eVALUate* is open via the student portal and each week non-responders are sent reminder emails encouraging them to give feedback. Teaching staff also encourage students to give feedback and posters are displayed across campus. In the two main evaluation periods, *eVALUate* is open for student feedback for a six week duration (during the last three weeks of class, through study week and the two week examination period). The system captures the date and time of survey submission; which is converted into the week of survey submission.

This study examined the *eVALUate* data gathered in four consecutive major evaluation periods (Semesters 1 and 2 in both 2009 and 2010). Survey responses were analysed to determine: 1) differences in response rates over time for different student subgroups, and 2) effect of timing of survey submission. The percentage Agreement (percentage of responses with Agree or Strongly Agree) for each quantitative survey item was calculated for each of the six weeks in each semester at the university and faculty levels. Chi square was calculated to test significance between student responses and week of survey submission and Cramer's V was calculated to control for the large sample size and to measure effect size.

Results

A total of 176,134 survey responses out of a possible 466,431 surveys were submitted over the four semesters. As shown in Table 1, a higher proportion of responses were submitted in weeks one to three and six (between 17.3% and 19.8%) than in weeks four and five (13.7% and 12.2% respectively).

Table 1. Percentage of responses submitted by week of submission

Survey submission week	Sem 1, 2009		Sem 2, 2009		Sem 1, 2010		Sem 2, 2010		All 4 semesters	
	Resp	% of resp	Resp	% of resp	Resp	% of resp	Resp	% of resp	Resp	% of resp
Week 1	8573	18.8%	6713	17.0%	10690	22.0%	8817	20.8%	34793	19.8%
Week 2	8898	19.5%	6185	15.6%	8818	18.2%	6528	15.4%	30429	17.3%
Week 3	9989	21.9%	7580	19.2%	8695	17.9%	7793	18.3%	34057	19.3%
Week 4	5568	12.2%	5313	13.4%	6103	12.6%	7142	16.8%	24126	13.7%
Week 5	5060	11.1%	5320	13.4%	5957	12.3%	5202	12.2%	21539	12.2%
Week 6	7487	16.4%	8446	21.4%	8266	17.0%	6998	16.5%	31197	17.7%
Total	45575		39557		48529		42480		176141	

Note: Sem = Semester; Resp = responses

A greater proportion of responses over the four semesters were submitted in week six in both the Faculty of Science and Engineering and the Faculty of Health Sciences. There was little variation in weekly response rates over the six weeks by age group, gender or semester weighted average, although there were less responses in week six by students of low semester weighted average (14.0% of responses by students with less than 60% semester weighted average were submitted in week 6 compared to 19.0% of responses by students with 60% or higher semester weighted average).

Differences in student responses by week of survey submission varied over the two years. In both semesters in 2009, responses were generally lowest in week one (third last week of class). In both semesters in 2010, responses were generally lowest in week five (first week of exams). In all four semesters, responses were generally highest in week three (last week of class) and four (study week), except for the student motivation and engagement items, in which responses were highest in week six (final week of exams) in all three semesters except Semester 1 2010 when they were highest in week five (first week of exams).

Differences in survey results by week of survey submission were relatively small. In the 11 quantitative items over the four semesters, the difference in percentage Agreement between the week with the highest percentage Agreement and the week with the lowest percentage Agreement ranged from 0.8% to 5.1%. The mean difference between the week with the highest level and the week with the lowest level of percentage Agreement in the survey items over the four semesters was 2.9%. The survey item with the greatest variation between lowest and highest

week over the semesters was Item 5 (Feedback) and the survey item with the least variation was Item 1 (Outcomes). Differences between highest scoring week and the lowest scoring week were lowest in Semester 1 2010, when they were below 2.0% in seven of the 11 items and below 3.0% in the other four items. Differences were greatest in Semester 2 2009 when they were greater than 3.0% in two thirds of the items.

Table 2 shows the results by week of survey submission. The week with the highest percentage Agreement per semester for each item is highlighted in light grey while the week with the lowest percentage Agreement per semester for each item is highlighted in dark grey. Difference refers to the difference in percentage Agreement between the week with the highest percentage Agreement and the week with the lowest percentage Agreement. As the table shows, results over the six weeks were relatively stable over different semesters and across the different weeks, even in the items with the greatest variation over survey submission week (the Feedback, Quality of teaching, and Overall satisfaction Items).

Table 2. eVALUate overall results per semester by week of survey submission

	Week of survey submission	Outcomes	Experiences	Resources	Assessment	Feedback	Workload	Teaching	Motivation	Best use	Think about	Overall satisfaction
S1 2009	Week 1	86.8	81.6	81.4	82.3	73.0	82.1	79.1	82.8	83.8	81.7	81.0
	Week 2	88.4	84.4	83.4	83.7	77.6	83.8	82.0	84.4	83.7	82.7	83.3
	Week 3	88.6	84.7	84.4	84.9	78.1	84.9	83.1	84.2	84.3	83.3	84.2
	Week 4	88.1	84.8	84.1	83.7	76.5	84.2	83.0	84.3	83.5	82.1	83.1
	Week 5	87.6	83.0	82.0	83.6	75.2	84.2	81.2	84.0	84.3	84.0	82.8
	Week 6	87.6	83.1	83.1	82.9	76.3	83.5	82.2	85.2	85.5	85.0	83.1
	Overall	87.9	83.6	83.1	83.6	76.3	83.8	81.8	84.1	84.2	83.1	82.9
	Difference	1.8	3.2	3.0	2.6	5.1	2.8	4.0	2.4	1.8	3.3	3.2
S2 2009	Week 1	87.1	82.2	82.6	82.6	74.9	82.9	80.8	82.6	83.8	82.7	81.3
	Week 2	88.5	85.2	84.3	84.5	78.2	84.9	83.7	84.6	85.7	85.0	83.4
	Week 3	89.6	86.1	85.6	86.2	79.4	86.4	84.3	85.1	85.0	84.0	84.6
	Week 4	88.8	85.4	85.3	85.4	78.8	86.8	84.5	85.3	85.4	83.2	84.2
	Week 5	87.3	82.6	82.0	82.5	75.3	83.6	81.7	83.4	84.0	83.2	81.7
	Week 6	88.4	84.5	84.0	83.6	76.5	85.1	83.2	85.5	86.1	85.1	83.6
	Overall	88.3	84.4	84.0	84.2	77.2	85.0	83.0	84.5	85.1	83.9	83.2
	Difference	2.5	3.9	3.6	3.7	4.5	3.9	3.7	2.9	2.3	2.4	3.3
S1 2010	Week 1	87.9	83.4	83.4	84.3	77.1	84.4	81.5	84.6	85.4	83.7	82.8
	Week 2	88.4	84.6	85.1	84.6	78.5	84.7	83.2	85.2	85.8	84.0	83.7
	Week 3	88.7	85.4	85.0	85.2	78.7	85.5	84.0	84.9	85.9	84.0	84.4
	Week 4	88.1	85.2	84.3	84.4	77.8	85.1	82.4	84.2	85.2	83.2	83.2
	Week 5	88.3	83.3	83.8	83.3	75.8	83.7	82.1	84.5	84.9	83.9	82.1
	Week 6	88.2	84.5	84.2	83.5	77.3	84.3	83.6	85.7	86.1	84.9	83.5
	Overall	88.3	84.4	84.3	84.3	77.6	84.6	82.8	84.9	85.6	84.0	83.3
	Difference	0.8	2.1	1.7	1.9	2.9	1.8	2.5	1.5	1.2	1.7	2.3
S2 2010	Week 1	87.8	83.9	83.6	84.2	77.7	85.4	83.2	83.9	85.3	82.9	83.0
	Week 2	88.9	85.8	85.5	85.5	80.5	85.7	84.3	85.0	85.2	84.1	85.1

Week 3	89.5	86.5	86.6	86.5	81.4	86.3	85.1	85.3	85.4	84.1	85.8
Week 4	88.9	85.2	84.5	84.8	79.6	86.8	84.1	84.7	85.3	84.3	84.7
Week 5	87.4	83.8	83.5	83.3	76.3	83.5	81.3	84.1	84.5	84.7	82.4
Week 6	88.3	86.1	85.7	85.4	78.3	86.1	83.8	86.5	87.2	86.2	85.1
Overall	88.5	85.2	84.9	85.0	79.1	85.7	83.7	84.9	85.5	84.3	84.4
Difference	2.1	2.7	3.1	3.2	5.1	3.3	3.8	2.6	2.7	3.3	3.4

Note: S = Semester

Chi Square values based on student feedback by week of submission were significant in almost all survey items in each semester. Of the 11 items over four semesters, there were only two instances when the Chi Square was not significant. Cramer's V values varied over the different semesters and across different survey items. In Semester 1 2009, the effect size was weak for about half of the survey items and moderate to strong for the other items. In Semester 2 2009, the effect size was moderately strong or strong in most items. In Semester 1 2010, the effect size was weak or insignificant in all items apart from two in which there was a moderate effect. In Semester 2 2010, there was a moderate effect in about half the items and a moderately strong or strong effect in four of the items. Only two survey items were found to have a moderate or moderately strong effect across all four semesters and these were Item 7 (Quality of teaching) and Item 11 (Overall satisfaction). Item 5 (Feedback) had a strong association in three of the four semesters but only a weak association in the other semester.

Figures 1 to 3 show the results by week of submission for each of the four semesters in three of the eVALUate unit survey items.

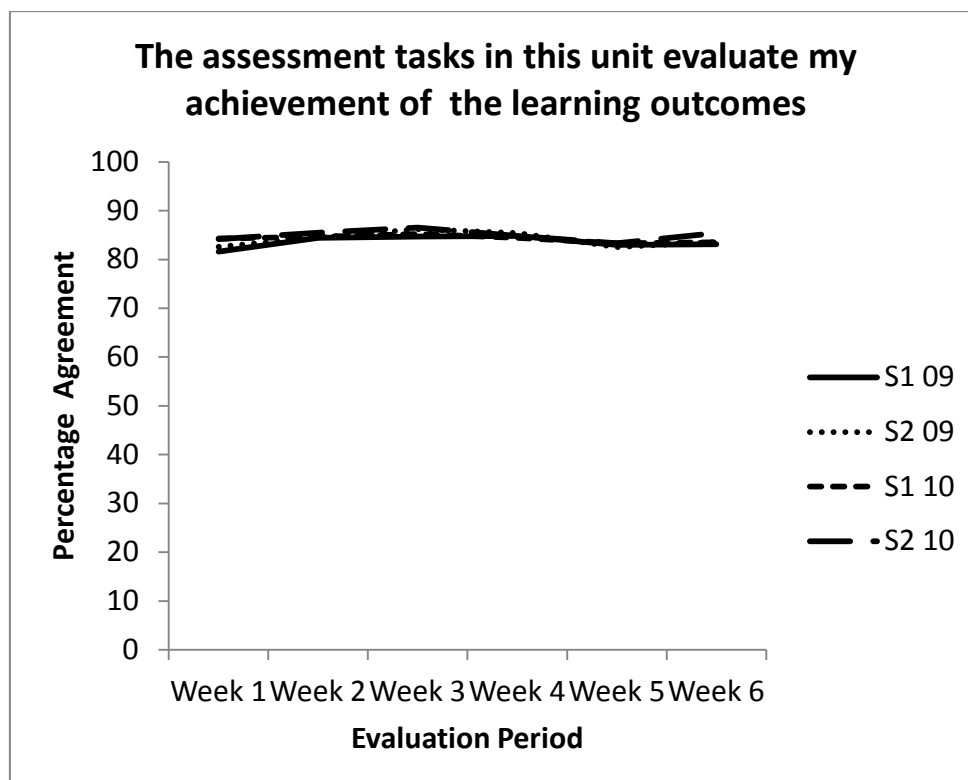


Figure 1. eVALUate results by week over four semesters in Item 4 (Assessment)

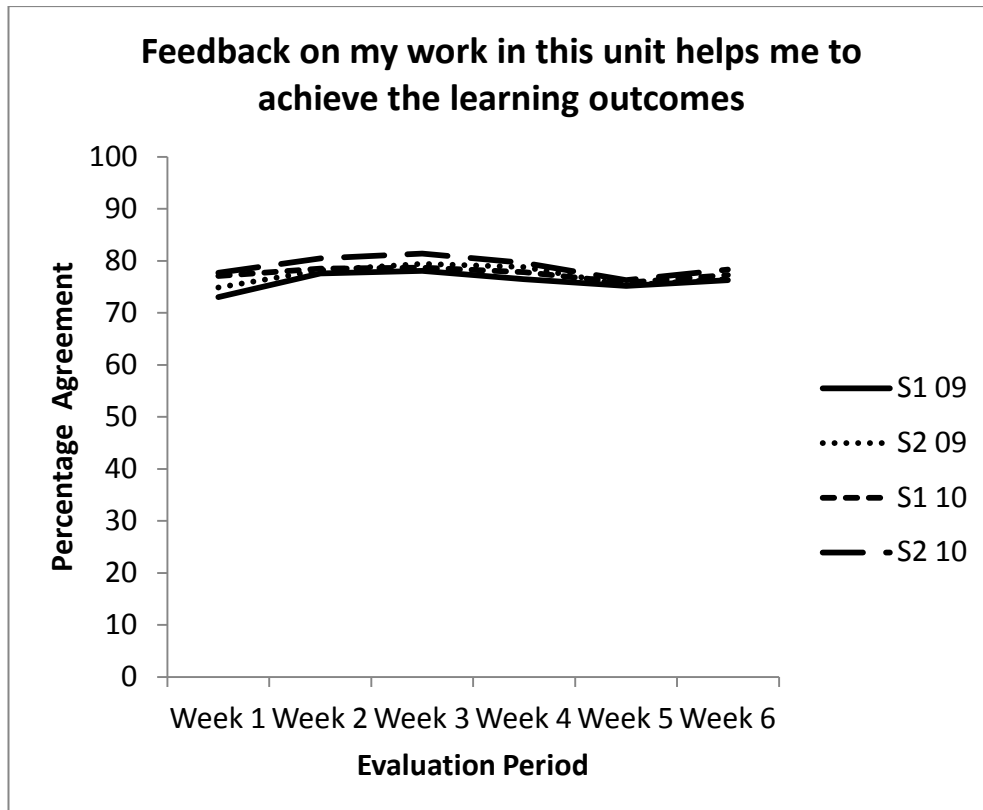


Figure 2. eVALUate results by week over four semesters in Item 5 (Feedback)

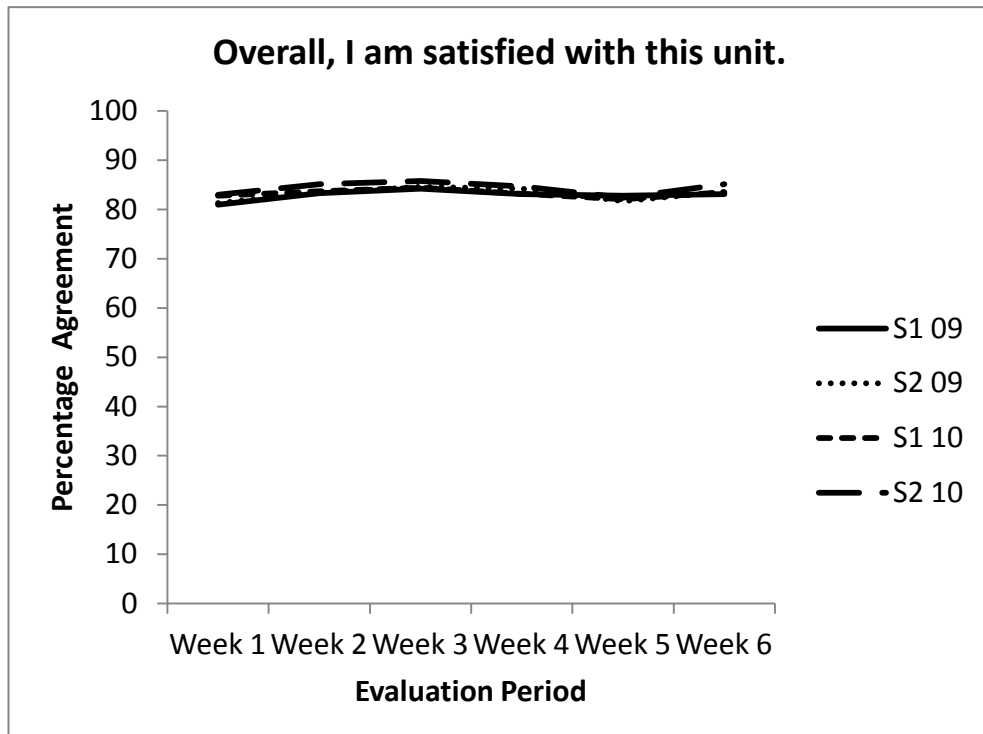


Figure 3. eVALUate results by week over four semesters in Item 11 (Overall Satisfaction)

In the Faculty of Humanities, percentage agreement for all 11 items was nearly always at its lowest level in week one. Results were at their highest in different

weeks over the four semesters. There were greater differences between percentage Agreement in the highest scoring week and lowest scoring week in the Faculty of Humanities than there were in the University overall. Differences were up to 8.6% in Humanities compared to differences up to 5.1% in the University overall. Only about half of the differences were below 5.0% in Humanities. The survey item with the most variation across week of submission in the four semesters was Item 11 (Overall satisfaction), followed closely by Item 10 (Think about).

In the Curtin Business School, percentage Agreement for all 11 items was at its lowest level in week one in Semester 1 2009, however it was generally at its lowest level in week five in the other semesters. Results were at their highest level in weeks two, three or four in all semesters. Differences in results between highest and lowest scoring weeks in the Faculty were greater than in the University overall and the other Faculties. Differences were up to 10.2%. There was an increase in differences in the Faculty from 2009 to 2010. The survey item with the most variation over the weeks was Item 5 (Feedback) while the item with the least variation was Item 10 (Think about learning).

In the Faculty of Science and Engineering, percentage Agreement was generally lowest in week one but for some items in some semesters, it was lowest in week two or week three. Results were at their highest in different weeks for different items across the different semesters, except in Semester 1 2010 when results were at their highest in week four for almost all survey items. Differences between highest and lowest scoring weeks in the Faculty were greater than in the University overall. Differences were up to 7.2%, although the majority of differences were below 5% and there were smaller differences in 2010 than in 2009. The survey item with the most variation was Item 5 (Feedback) and the item with the least variation was Item 1 (Outcomes).

In the Faculty of Health Sciences, results were almost always at their lowest in week one. In 2009, results were generally highest in week three or six whilst in 2010 results were generally highest in week six. There were smaller differences between lowest scoring week and highest scoring week in this Faculty than in the other Faculties. Differences were up to 6.6% but the majority of differences were below 5%. The survey item with the most variation across weeks in the four semesters was Item 5 (Feedback) and the item with the least variation across the weeks was Item 1 (Outcomes).

Discussion

A greater number of surveys were submitted in the first three weeks and the final week of the six week evaluation period. Lower response rates in later weeks are expected since there is a smaller remaining population of non-responders in each successive evaluation week. The reduction in survey submissions in weeks four and five may also be a reflection of student activity in these weeks: week four is the study week prior to exams and week five is the first examination week. Many students are still undertaking exams in week six, and with encouragement from academic staff, may be more likely to submit their surveys after completing all their exams.

Consistent with the findings of Coffey and Gibbs (2000), students with a low semester weighted average were less likely than students of higher semester weighted average to give feedback after their exams.

A greater proportion of responses over the four semesters were submitted in week six in both the Faculty of Science and Engineering and the Faculty of Health Sciences. This is possibly due to students in those Faculties finishing all their assessments later than students in the other Faculties.

Survey results varied slightly over the weeks and generally were lowest in weeks one and five and highest in weeks three and four. The differences between the week with the lowest percentage Agreement and the week with the highest percentage Agreement were small, ranging from less than 1% to just over 5%. Whilst Cramer's V values showed there were moderately strong and strong effect sizes in some items (Quality of Teaching, Feedback and Overall Satisfaction) in several semesters, this was not a consistent trend. In one semester, there was a weak effect or no effect in the majority of survey items. What is notable is the stability of the student feedback over the six weeks, particularly considering the many inherent factors which potentially impact on student feedback. These include: the unique assessment patterns and weighting of end of semester examinations; the different disciplines; subject topics; learning experiences; variations in teaching methods and teaching styles.

There may be extrinsic factors which could potentially impact on the feedback including differing levels of student stress over the six weeks. With responses at their lowest in weeks one and five (third last week of class and first week of exams) and highest in weeks three and four (last week of class and study week), it is conceivable that student evaluation results are impacted by student stress in the run up to the examination period. However, this explanation does not account for the higher results of the study week (week 5). It may be that student feedback is more positive when students are more positive because their classes are drawing to a close or finished (weeks three and four) and their exams are drawing to a close or are completed (week six) or because the fear of the unknown (with respect to final grades and/or exam difficulty) is over, as suggested by Leigle and McDonald (2005).

Conclusions

This study investigated differences in eVALUate participation rates over the six week evaluation period in four major student evaluation periods (semesters). The study also examined differences in survey responses over the six weeks. There was no consistent pattern of variation over weeks of submission across the four different semesters. There were instances of strong and moderately strong effect size for week of survey submission but for the most part, the survey responses were stable over time. Differences were generally small and whilst there were some recurring patterns of variation in responses, there were also quite varied patterns of responses over different semesters in different weeks, particularly at the faculty level. The overall stability of survey responses over time provide reassurance to academic staff about promoting eVALUate at specific times during the evaluation period, including after examinations.

References

- Coffey, M., & Gibbs, G. (2000). Can Academics Benefit from Training? Some Preliminary Evidence. *Teaching in Higher Education*, 5(3), 385-389.

- Gillard, J. (2010). *Government to introduce 'My University' website*. Education, Employment and Workplace Relations portfolio, Ministers' Media Centre, 3 March, Retrieved from http://www.deewr.gov.au/Ministers/Gillard/Media/Releases/Pages/Article_100303_153420.aspx
- Krantz-Girod, C., Bonvin, R., Lanares, J., Cuénot, S., Feihl, F., Bosman, F., & Waeber, B. (2004). Stability of repeated student evaluations of teaching in the second preclinical year of a medical curriculum. *Assessment and Evaluation in Higher Education*, 29(1), 123-133.
- Leigle, J., & McDonald, D. (2005). Lessons Learned From Online vs. Paper-based Computer Information Students' Evaluation System. *Information Systems Education Journal*, 3(37).
- Marsh, H.W. (2007). Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness. In RP Perry & JC Smart (Eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, pp 319-374. Dordrecht, The Netherlands: Springer.
- Marsh, H.W., & Hocevar, D. (1991). Students' evaluations of teaching effectiveness; the stability of mean ratings of the same teachers over a 13-year periods. *Teaching and Teacher Education*, 7(4), 303-314.
- McNulty, J.A., Gruener, G., Chandrasekhar, A., Espiritu, B., Hoyt, A., & Ensminger, D. (2010). Are online student evaluations of faculty influenced by the timing of evaluations? *Advances in Physiology Education*, 34, 213-216.
- Oliver, B., Tucker, B., Gupta, R., & Yeo, S. (2008). eVALUate: An evaluation instrument for measuring students' perceptions of their engagement and learning outcomes. *Assessment & Evaluation in Higher Education*, 33(3), 619-630.
- Oliver, B., Tucker, B., & Pegden, J. (2007). An investigation into student comment behaviours: Who comments, what do they say, and do anonymous students behave badly? In conference proceedings *Australian Universities Quality Forum 2007*. (Hobart, Tasmania, 12-13 July). Retrieved 23 March, 2012, from http://pandora.nla.gov.au/pan/127066/20110530-1318/www.auqa.edu.au/files/publications/auqf2007_proceedings_final_website_text.pdf

Appendix

The eVALUate Unit Survey

Quantitative items with the following rating scale (Strongly agree, Agree, Disagree, Strongly disagree and Unable to judge). *Explanatory text in italics appears online by default.*

1. The learning outcomes in this unit are clearly identified.
The learning outcomes are what you are expected to know, understand or be able to do in order to be successful in this unit.
2. The learning experiences in this unit help me to achieve the learning outcomes.
The learning experiences could include: face-to-face lectures, tutorials, laboratories, clinical practicums, fieldwork, directed learning tasks, and online learning experiences.
3. The learning resources in this unit help me to achieve the learning outcomes.
Learning resources could include print, multimedia and online study materials, and equipment available in lectures, laboratories, clinics or studios.
4. The assessment tasks in this unit evaluate my achievement of the learning outcomes.
Assessment tasks are those which are rewarded by marks, grades or feedback. Assessment tasks directly assess your achievement of the learning outcomes.
5. Feedback on my work in this unit helps me to achieve the learning outcomes.
Feedback includes written or verbal comments on your work.
6. The workload in this unit is appropriate to the achievement of the learning outcomes.
Workload includes class attendance, reading, researching, group activities and assessment tasks.
7. The quality of teaching in this unit helps me to achieve the learning outcomes.
Quality teaching occurs when knowledgeable and enthusiastic teaching staff interact positively with students in well-organised teaching and learning experiences.
8. I am motivated to achieve the learning outcomes in this unit.
Being motivated means having the desire or drive to learn, to complete tasks and to willingly strive for goals.
9. I make best use of the learning experiences in this unit.
I prepare for and follow up on the learning experiences offered in this unit.
10. I think about how I can learn more effectively in this unit.
I take time to think about how I can learn more effectively.
11. Overall, I am satisfied with this unit.
Overall, this unit provides a quality learning experience.

Qualitative items

- 12.* What are the most helpful aspects of this unit?
- 13.* How do you think this unit might be improved?