IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITP4385

A Critical Reflection on the Value of Course Evaluation Questionnaires: Using Group Support Tools to Explore Student Evaluations

David A Banks
School of Accounting and Information Systems, University of South Australia; david.banks@unisa.edu.au

ABSTRACT

End of course evaluations by students can, in combination with the lecturers own perception of the outcomes, provide staff with useful feedback that can guide future deliveries of the subject. However, the formalisation of course evaluation questionnaires and their use for assessment of staff requires that the validity of such instruments is carefully considered. This paper suggests that the evaluation of outputs from such processes needs to be considered in the light of the variability of students interpretation of the questions before any value can be obtained. Two approaches using keypad and keyboard group support technology to explore student evaluations are outlined and the outcomes reflected upon.

INTRODUCTION

The use of student evaluation questionnaires is well established practice in many academic institutions and would appear to be a potentially useful tool to help members of staff review their teaching and make any appropriate changes to future repeats of the subject. The results of the feedback will normally be in addition to the day to day feedback given informally by students and to the broader perceptions of the staff member. However, there is a move in some institutions to use the raw data derived from simple questionnaires as a formal vehicle for assessing the teaching performance of staff and this may be a cause for some concern. The quantitative interpretation of the data without some understanding of qualitative issues of teaching style, process, environment and so on would seem to be a rather minimalist approach to a process that could seriously disrupt the career or teaching style of the member of staff. The use of a teaching style, for example that is radically different to the norm could generate adverse raw scores even though the approach is educationally sound.

My own experiences with instruments that generate a simple numerical 'measure' of students impressions of such areas as 'overall quality' or 'performance of the lecturer' have left me feeling rather nervous about their validity. The presentation of a list of questions without explanation or discussion leads to the situation where students are free to interpret the meaning of the questions from a variety of standpoints. The 'input' part of the process is therefore subjective on the part of the student and provides an interpretative or qualitative view of the world driven by factors of language, context, framing etc. Any attempts to interpret the resulting data on a purely quantitative, or positivist, basis would appear to be somewhat flawed and simplistic as an approach.

In 1994 I facilitated a one-day intensive session on the Henley Management Programme in New Zealand, in which, given the nature of the participants, I deliberately did not use formal lectures but instead used a mix of workshops, discussion, cases, exercise etc. The results of the subsequent paper-based evaluation were somewhat unhelpful to me as one student scored the 'lectures' item as 'excellent', but another student indicated on the evaluation form that I did not use lectures. If students cannot agree on what a 'lecture' is then I have doubts about the validity of any attempts to interpret the raw score data. On further reflection I wondered if it is possible to score a maximum for some items, (there may be resource constraints outside my control) or what efforts or actions are required to move an average score from, say, 3.1 to 3.5, or if a move from 3.1 to 3.6 is the same as a move from 3.6 to 4.1.

On return to my UK university I was greeted by a new student evaluation form and an indication that explanations would be required if the scores fell below a specified numeric value. When questioned about what the significance of the chosen 'cut-off' level indicated it became clear that the 'Quality Manager' had little rational basis for its selection.

As a result of these concerns about evaluation in general I carried out some work using a portable keypad based group support system (OptionFinder) with a small group of nine students taught by another member of staff (the Quality Manager in fact) within the School of Engineering and IT. The member of staff who taught the students was not present at the meeting but a non-teaching member of staff was present as an observer.

THE OPTIONFINDER SESSION

OptionFinder is a group support tool that allows group members to enter anonymous data through a hand-held keypad similar to a television remote control. Questions to be explored are entered into the software prior to the meeting and during the meeting data from the keypads is gathered (ie polled) anonymously and displayed in various formats on a public screen via a data projector. For the first poll the students were asked to use the keypad to individually score each item as it appeared on the public screen, without making any comments. After all items had been scored the results from the poll were displayed and discussion took place with the students in an effort to try to understand the possible reasons for the pattern of distribution of votes. This involved discussing the criteria that students used when they were scoring each item. After this discussion the students were asked to vote again on each item using their individual keypads. The items and average scores for the first and second poll data are shown below:

Evaluation topic	First poll average	Second poll average
Support for independent study	2.7	2.1
Quality of course materials	3.3	3.1
Extent to which learning outcomes are met	3.0	3.0
Helpfulness of tutors comments	2.4	2.6
Standard of teaching	3.4	3.0
Clarity of assessment methods and marking criteria	3.4	3.2
Quality of feedback	2.7	2.4
Overall satisfaction with the unit	3.2	3.1

This paper appears in *Managing Information Technology in a Global Economy*, the proceedings of the Information Resources Management Association International Conference.

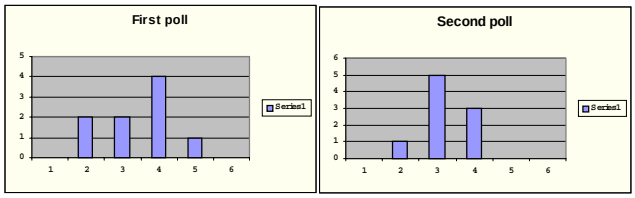
(A six point scale was used, in line with the paper questionnaire, with 1 indicating Very Dissatisfied and 6 indicating Very Satisfied).

The above numeric outcomes of the meeting do not offer much insight, except perhaps to suggest that the students have, in some instances, become more critical as a result of sharing experiences and views. Further consideration of the detailed voting patterns before and after discussion does, however, reinforce the idea that consensus formation had taken place:

	First poll							Second poll					
(Score)	1	2	3	4	5	6		1	2	3	4	5	6
Support for independent study	1	3	3	2	0	0		1	6	2	0	0	0
Quality of course materials	0	2	3	3	1	0		0	0	1	6	2	0
Extent to which learning outcomes are met	1	1	4	3	0	0		0	1	7	1	0	0
Helpfulness of tutors comments	1	3	5	0	0	0		1	2	6	0	0	0
Standard of teaching	0	0	6	2	1	0		0	2	5	2	0	0
Clarity of assessment methods and marking criteria	0	2	2	4	1	0		0	1	5	3	0	0
Quality of feedback	1	4	1	3	0	0		0	5	4	0	0	0
Overall satisfaction with the unit	1	0	4	4	0	0		0	2	4	3	0	0

(The numbers in the cells show the number of students who voted for a particular scale item)

The numeric changes between polls for each item are quite close, for example 'Clarity of assessment methods' moves from 3.4 to 3.2 and one would be hard pressed to derive much information from this. However an examination of the shape of the distribution does suggest some insights:



In the first distribution there is a fairly broad spread of voting pattern whereas the second distribution would seem to indicate a move towards closer agreement after the basis for the first voting pattern had been discussed. The shape of the distribution is significant and needs to be taken into account in any interpretations. The obvious danger is that a group that totally agrees on a middle level score would generate the same score as a bipolar vote pattern. The raw number conveys little useful information but the overall shape could suggest lines of investigation to determine why, for example, the group had polarised or broadly distributed views.

Discussion with the students indicated that their main concerns, which strongly influenced their voting patterns, were related to general resourcing, specific facilities such as computer provision, library etc rather than directly to the subject under evaluation. 'Resources' was also taken to include the Cocoa Cola machine, which had proved to be somewhat unreliable that semester. It became evident that the students did not base their evaluations of the subject on just the subject itself, but upon the whole of their experiences within the university for that term. The major factors that influenced the students levels of satisfaction with the actual subject under consideration, rather than their whole university experience, transpired to be the duration of the sessions (too long without a break) and the use of inappropriate furniture in the room. The introduction of a break into the session and a change of furniture had considerable positive impact on the student group. Clearly the raw numbers would have done little to reveal this simple, and cheap, way to improve student perceptions and the students indicated that they did not feel that such operational issues fell within the remit of the questionnaire. They felt that the paper questionnaire was of little practical concern to them and indicated that the forms were completed with little thought as to the value of the scales being used. They also felt that it was impossible to separate the appraisal of a subject from their total experience of the university and that this would lead to a position where some items on the list could not be scored as 5 despite the best efforts of the member of staff. The session proved to be successful, running beyond the one hour initially planned. The students felt that this was the first occasion in their university careers that they had been given an opportunity to explore issues to any depth and strongly supported the use of the keypad based approach. The process of discussing the data and being able to express and share ideas was felt to be extremely valuable.

A TEXT BASED SESSION USING GROUPSYSTEMS

Upon moving to the University of South Australia I gained access to GroupSystems software, based in the EPICentre in the School of Accounting and Information Systems, and this offered the opportunity to further explore the problems of interpretation of evaluation data gathered from student groups. GroupSystems allows textual data from a number of terminals to be gathered and displayed on a public screen. Once again the input is anonymous. In an effort to try to understand what criteria students used when completing course evaluation questionnaires (CEQ) students studying Information Systems Policy (a final year subject delivered by the author) were asked to complete the CEQ process using GroupSystems. It was felt that this would be a useful subject to focus upon as it had been taught from a very 'soft' perspective of information systems in contrast to the 'harder' leaning of most of their previous subjects. Sixty-six students were involved in this exercise which comprised five one-hour sessions with approximately fourteen students in each session. The facilitation of the session was carried out by myself with the support of a technographer (who operates the software) who also knew some of the students.

The procedure adopted was to provide the students with the normal hard copy of the CEQ and ask them to score in the usual way using a scale of 5, with 1 indicating 'Strongly Disagree' and 5 indicating 'Strongly Agree'. Having scored all items they then used the GroupSystems Categoriser tool to open the 'bucket' that was labelled with the score they had marked on the hard copy and to enter their reason(s) for scoring that particular value. The time constraints for this process dictated that a subset of six questions was selected from the normal twenty-five CEQ. These were chosen to explore areas that had been indicated to me as being of some concern to students in previous deliveries of the subject. The questions were:

The teaching staff of this subject motivated me to do my best work
It was often hard to discover what was expected of me in this subject
The lecturer was extremely good at explaining things
The teaching staff worked hard to make this subject interesting
The teaching staff made it clear right from the start what they expected from students
Overall I was satisfied with the quality of this subject

Results for 'Overall I was satisfied with the quality of this subject'

Total responses per score:

Score	Responses
1	9
2	5
3	26
4	19
5	7

The questionnaire form clearly indicated that a score of 1 equated to "Strongly Agree" and a score of 5 to "Strongly Disagree". One would therefore anticipate that the entries in the "Score of 5" bucket would indicate why the student was unhappy with the quality of the subject. Actual comments in the 'Score of 5' bucket were:

1. What a crock - I gave it a five
2. I was satisfied with the level of the teaching
3. I have respect for anyone who stands in front of a class and asks for anonymous critique!
4. Good mix of assessment methods, lecturer always accessible.
5. this was a quality ARTS subject, I felt that it was not orient towards the business degree enough and as such did not constitute high quality
6. quality yes, and especially that he didn't demand too much. so quality for time both put in and required.
7. Yah mate! you are great but please make sure i pass the subject

The comments above do not all appear to be in the anticipated bucket, particularly items 2 and 4.

Looking at the other end of the scale, (score of 1, ie Strongly agree that 'Overall I was satisfied with the quality of this subject') where one would expect that students had some positive reasons for satisfaction reveals:

1. Excellent group sessions, formal and informal, and techniques for communication
2. the overall quality of this subject was good, i have to say its been my favourite subject, work was not to hard, teacher/ tutor/lecturer was good, web paged helped a lot
3. The subject entailed the debate and a major paper which really made you think about issues (especially in debate). Thumbs up.
4. No exams, oh boy what a way to go!!!
5. was happy with the subject and its content as to how it varied but was something that I could not feel passionate about
6. satisfied because I improved on my way of researching process
7. the subject had quantity don't know about quality
8. depends if i pass or not - but otherwise it was directionless and pointless, didn't learn much just did the work, same as ecommerce. David tried to over intellectualise an essentially unintellectual degree
9. I am rather satisfied with the whole subject. I think it will be a bonus to get some specific help on the issues of policy making pertaining to the major assignment worth 60%. This is really a killer to achieve a good grade in this subject. It will be much appreciated if this area can be further improved on the part of the lecturer

Once again one would need to ask if items 7 and 8 are in the 'correct' bucket.

The middle ground, a score of '3' indicated a wide variety of comments:

I am still not sure of this subject. Policy is so undefined. It is very difficult to feel comfortable with this subject, therefore it is difficult to comment on quality. I did feel it lacked some structure though.

A little more guidance would have helped. The fact that there was no exam for this subject in third year was pretty cool

When I think about quality I think about how I liked the subject overall. I had mixed feelings, therefore I marked a 3. I like the lecturer and the assessment methods (LOVE no exam), but not the subject matter.

I have studied better subjects and worse subjects.

This subject is worth studying because it provides the ideas of what is a policy is and using the EMS as new experience

I enjoyed the subject and the assignments were varied and interesting but I am unsure of the relationship of the subject to others in the course.

It would have been nice to have more time on assessments

No exam, can't complain!

It was certainly different to the more traditional subjects; it promoted self learning which I think was important in the overall context

If somebody can define what I perceive as quality and what aspect of the subject I should prioritize in creating an overall score other than 3 then they are far better than me

The subject evaluation questionnaire needs a bit of work

The outside the square thinking is GREAT!!!! The fact that the final paper is worth 60% is not so great...

Quality depends on how much work individuals put in.

In the middle of the road I sit. Quality? How do you define quality? It provided a broad range of subject matter (that one day I will understand somehow fits into IS policy).

Yes I was satisfied, there was no pressure and you could pretty much do things at your own pace

It was okay - got to do your own thing - and there is no exam

Yes, in the end it was OK! Lets just hope we all pass!

Many of the above comments warranted deeper discussion but time constraints meant that only superficial exploration was possible. The lack of exam seems to have been seen as a positive aspect, but in informal discussion with some students they expressed the feeling that a final year subject should have an exam.

REFLECTION

The use of CEQ type instruments is widespread and may be seen as a way of assessing the performance of staff. However, attempts to base decisions about the performance of staff on simple numeric interpretation derived from an instrument that is passed to students without comment may be open to question. Most evaluation questionnaires include a comments section but these are often left blank and are also difficult to interpret without surrounding context. The keypad sessions indicated that a poll/re-poll approach can generate some shared understanding with a consequential likelihood that the second set of figures are probably more useful. The process of group discussion and identification of criteria to consider when evaluating educational experiences were also valuable for both staff and students. The GroupSystems sessions showed that students have a wide range of criteria in assessing the subject, not all of which may be those anticipated. There is some evidence that students were confused about the scoring system with positive comments appearing and the negative end of the scaling and vice versa.

An area worthy of further exploration is that of the value that individual members of staff place on the various questions within the questionnaire. As part of a staff workshop at the UK university we looked at the questionnaire and staff were asked to indicate which questions they felt were most valuable in helping them improve their future teaching. The keypad system was used to determine their rankings and a general discussion took place. Subsequent to the workshop I printed out the data based on the demographics of the School, which was an amalgamation of engineers (mainly electronic, communications and aeronautics), computing staff and staff from the old School of Systems and Information Sciences. The pattern of voting is shown below and may

suggest differing value systems were at play:

Engineering staff	*Computing staff*	*IS staff*
Quality of course materials	Helpfulness of tutor comments	Helpfulness of tutor comments
Helpfulness of tutor comments	Clarity of assessment	Quality of feedback
Clarity of assessment	Quality of feedback	Quality of course materials
Meeting learning outcomes	Meeting learning outcomes	Support for independent study
Support for independent study	Overall satisfaction	Clarity of assessment
Quality of feedback	Standard of teaching	Standard of teaching
Standard of teaching	Quality of course materials	Meeting learning outcomes
Overall satisfaction	Support for independent study	Overall satisfaction

('Most useful' item is at the top of each column)

The patterns of distribution raise many interesting questions and suggest a number of avenues for further study. If a particular academic has a particular perception of the relative values of evaluation items then it may be that this needs to be taken into account as yet another variable in the interpretation of the raw data output from the evaluation process.

My facilitation of both keypad and keyboard based approaches was as neutral as I could manage, but clearly there is a potential opportunity to manipulate students and direct them in ways that would favourably alter their scores. (Of course it can be argued that the paper-based approach is also open to manipulation, particularly if the career of an individual is at stake). It is also difficult to build in sufficient time to most subjects to allow extensive, and discursive, use of group support tools to conduct evaluation and also to provide sufficient resources to carry out these approaches on a wide scale. The text-based GroupSystems approach offers rich data but is time consuming and my efforts are now leaning towards the use of a thirty keypad group support system that has been made available to me by Interactive Meetings and Learning (IML) in Brisbane to support further investigations in this area.

CONCLUSION

The danger of the use of a paper based (or web equivalent) evaluation system that tries to judge the performance of staff on simple numeric scores, or leads to the abandonment of subjects that consistently generate low scores is that staff may simply adopt teaching strategies that generate 'required' scores. Many of the comments from the GroupSystems session above pave the road to such strategies – remove exams, avoid any level of ambiguity, 'hold hands', work from a single text, do not demand too much and so on.

In addition to a process of defensive teaching to gain or maintain high CEQ type scores there is also a danger that staff will not try to experiment or innovate if this could lead to poor scores, particularly in a system that does not fully investigate the reason for those scores. There may be occasions where low scores need to be accepted where the educational process is valid but unfamiliar to students, or is out of line with their normal expectations. A low score may be an indicator that students are having a poor educational experience, but equally it indicate that they are being exposed to novelty or are being taken into broader areas of study. It would be a sad future for education if the result of over-reliance on simplistic quantitative data led staff to abandon risky teaching and simply deliver a safe, non-threatening and non-challenging experience to the students.

Clearly we do need instruments that help us to gauge our performance in a number of aspects of teaching and learning, to support performance related issues and, perhaps more importantly, to help staff develop consistent and reliable benchmarks against which to test changes to content and to delivery mode and process. The work outlined above suggests that there may be a need for a broad and rigorous examination of the interpretation of the instruments and of the data generated by current instruments if they are to be provide the true value that they could offer.

ACKNOWLEDGEMENTS

My thanks go to Option Technologies in the UK for their initial support and to IML in both the UK and Brisbane for their continued support in this work.