

Student evaluation of degree programmes: the use of Best-Worst Scaling

Introduction

Students play a role as stakeholders in higher education quality enhancement through their involvement in educational evaluation (Stukalina, 2012). The Course Experience Questionnaire (CEQ) (Ramsden, 1991; Ramsden and Entwistle, 1981) has been used as a quantitative tool in various countries to evaluate student experiences of higher education programmes. In the case of Australia, parts of the CEQ have recently been administered in conjunction with the newly developed University Experience Survey. The CEQ has two essential purposes. The first relates to internal university objectives with a formative perspective on quality assurance. The second purpose is external in nature with a view to benchmarking (e.g. in university choice guides for students) and as a performance or accountability tool for governments.

The CEQ has its limitations in terms of both of the above purposes. Tucker et al (2008), for instance, observe that the long delay in obtaining CEQ data does not make them very suitable for continuous quality improvements. Marsh et al. (2011), on the other hand, criticise the CEQ for its inability to differentiate between universities. An area of criticism which has not received much attention relates to the fundamental issue of the method by which the evaluation responses are elicited from graduates. The CEQ, like most quantitative evaluation instruments, employs rating scales for that purpose. As discussed in Huybers (2014), item-by-item ratings data may suffer from well-known drawbacks including response bias and scale point inequivalence. In response, Huybers (2014) puts forward the use of Best-Worst Scaling (BWS) (Finn and Louviere, 1992; Marley and Louviere, 2005) as an alternative approach to student evaluation responses. The BWS method is choice-based which attends to the above two drawbacks.

Ratings responses often produce small differences between items of interest which, in the case of the CEQ, implies minimal formative educational insight into areas of relative strength and weakness. Australia wide CEQ results of recent years confirm that there is very little differentiation between evaluation items within and across its constituent scales. Compared with other elicitation methods, including rating scales, BWS has been shown to produce enhanced discrimination as well as greater predictive validity although survey completion times are longer (Chrzan and Golovashkina, 2006). The purpose of this paper is to report of the use of the BWS approach as the elicitation method for a degree evaluation instrument.

Method and Data Collection

In the BWS approach (to be precise, the BWS Case 1 variant, see Flynn and Marley, 2014), an overall list of items of interest is broken down into a series of

experimentally designed comparison sets. Respondents indicate, in each comparison set, their choice of the highest ranked (e.g. 'best' or 'most important') and the lowest ranked (e.g. 'worst' or 'least important') item.

The study was carried out at a Business faculty at a European university with a problem-based learning approach. Student experience evaluation appears particularly relevant in such a student-centred teaching and learning environment, and the CEQ has been applied in that context previously (for instance, Lyon and Hendry, 2002; Shamsan and Syed, 2009).

In the study, responses were collected from a representative sample of recent graduates at Bachelor level (n=82) and Masters level (n=181) (overall response rate of 19%). An on-line survey was used which incorporated 13 comparison sets of four items. The items were taken from an overall list of 13 items which included the items from two key CEQ scales: the Good Teaching Scale and the Generic Skills Scale (with slight adjustments in wording).

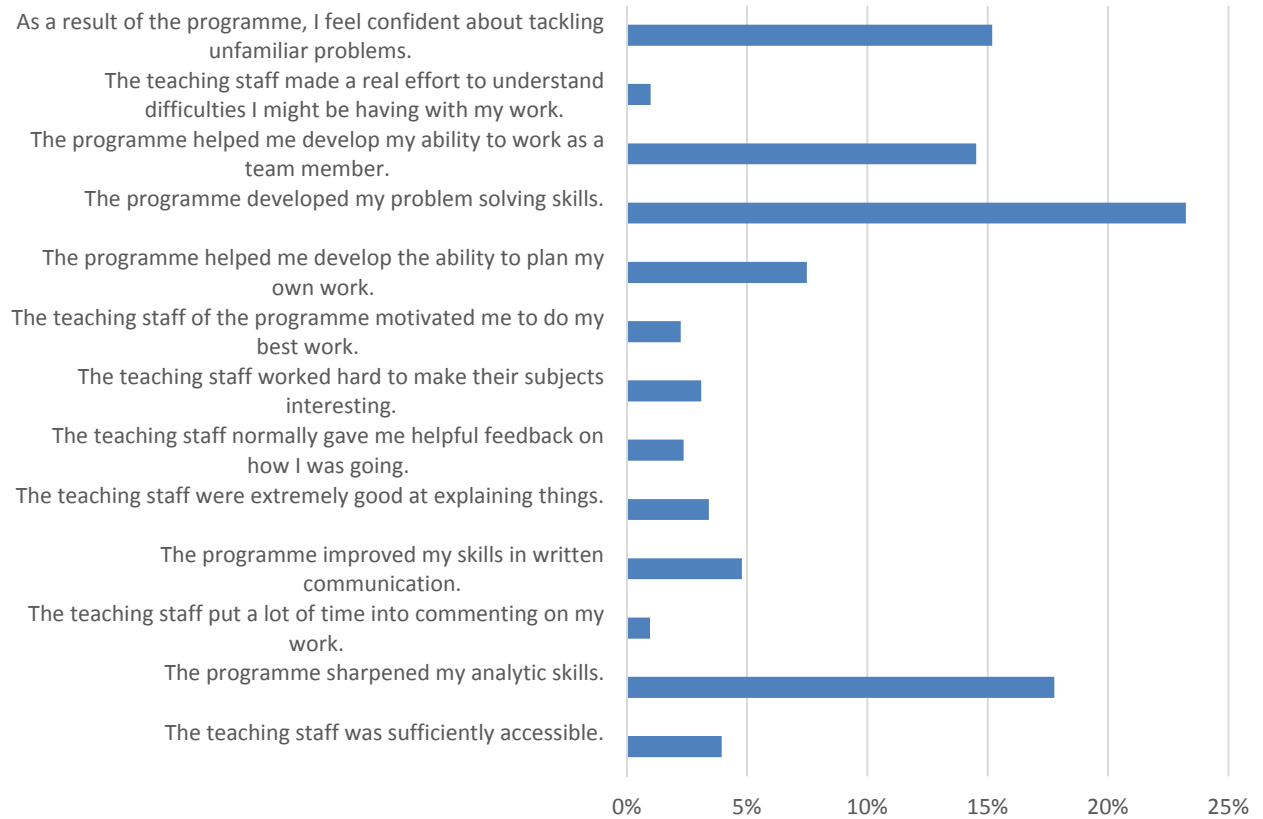
In each BWS comparison set, the graduates were asked to indicate the attribute that applied most to their degree programme and the attribute that applied least to their degree programme. In addition to the BWS experiment responses, the survey collected data on graduate characteristics including their degree, their programme, gender and country of origin.

Latent Class (LC) analysis was used to model the BWS data in conjunction with respondent characteristics. LC models produce, simultaneously, the optimal number of discrete classes in the sample, the parameters for each of the classes and class membership by way of respondent characteristics (Vermunt and Magidson, 2005; Wedel and Kamakura, 2000).

Findings and Discussion

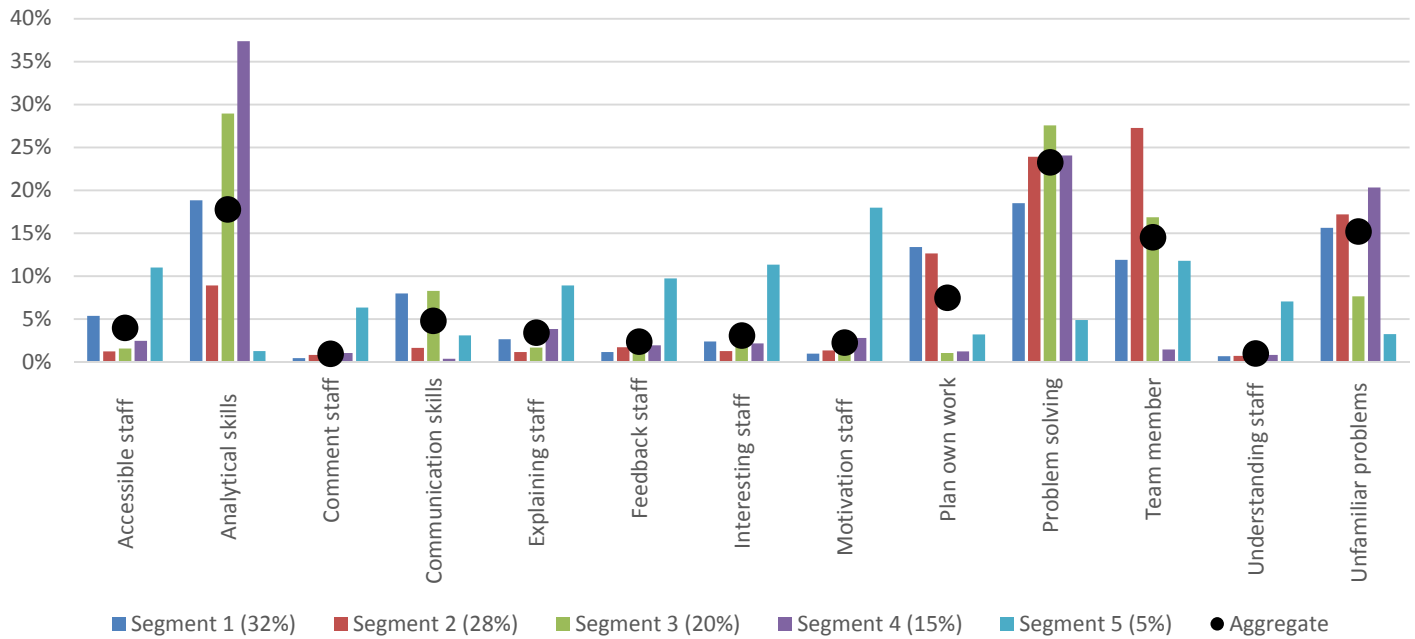
The LC model yields estimated parameters that are the utilities for each of the programme items evaluated. For ease of interpretation, the parameters can be converted into choice probabilities which add up to 100% for all 13 items. Figure 1 shows the whole-of-sample probabilities. The results show that the development of problem solving skills is the aspect that graduates associated most with their degree programme. This is followed by items relating to analytical skills, confidence in tackling unfamiliar problems and the ability to work as a team member. The choice probabilities for the other items, apart from work planning skills, are less than five percent which implies that those items were perceived as minor aspects of the degree programme. Overall, the results show that the skills-related items dominate the teaching-related ones. This appears to be consistent with Talukdar et al. (2013) who observe that the CEQ tends to focus on teaching and learning outcomes rather than the inputs or processes involved in achieving those outcomes.

Figure 1 Probability of the item being selected as most applicable, sample aggregate (n=263)



The LC analysis also identified, based on information criteria values, a model with five homogeneous preference classes. The choice probabilities for each of the segments (and for the sample aggregate as the reference case) are shown in Figure 2. The first four segments differ in the relative importance of the skills-related items while the fifth segment, the smallest in size, attached more relevance to teaching-related items.

Figure 2 Probability of the item being selected as most applicable, by segment



Overall, with a view to the purposes of the CEQ and similar programme evaluation instruments, the use of the BWS elicitation approach is promising with a view to establishing a greater degree of discrimination between the items evaluated. This would be expected to be of interest to university management in terms of quality insurance and improvement of their educational programmes.

References

- Chrzan, K., & Golovashkina, N. (2006). An empirical test of six stated importance measures. *International Journal of Market Research*, 48(6), 717–740.
- Finn, A., & Louviere, J. (1992). Determining the appropriate response to evidence of public concern: the case of food safety. *Journal of Public Policy & Marketing*, 11(1), 12–25.
- Flynn, T., & Marley, A. (2014). Best-worst scaling: theory and methods. In *Handbook of Choice Modelling*. Cheltenham: Edward Elgar Publishing, 178–201.
- Huybers, T. (2014). Student evaluation of teaching: the use of best–worst scaling. *Assessment & Evaluation in Higher Education*, 39(4), 496–513.
- Lyon, P., & G. Hendry (2002). The use of the Course Experience Questionnaire as a monitoring evaluation tool in a problem-based medical programme. *Assessment & Evaluation in Higher Education*, 27(4), 339–352.
- Marley, A., & Louviere, J. (2005). Some probabilistic models of best, worst, and best-worst choices. *Journal of Mathematical Psychology*, 49(6), 464–480.

- Marsh, H., Ginns, P., Morin, A., & Nagengast, B. (2011). Use of student ratings to benchmark universities: multilevel modeling of responses to the Australian Course Experience Questionnaire (CEQ). *Journal of Educational Psychology*, 103(3), 733-748.
- Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The course experience questionnaire. *Studies in Higher Education*, 16, 129–50.
- Ramsden, P., & Entwistle, N. (1981). Effects of academic departments on students' approaches to studying. *British Journal of Educational Psychology*, 51, 368–83.
- Shamsan, B. & Syed, A. (2009). Evaluation of problem based learning course at College of Medicine, Qassim University, Saudi Arabia. *International Journal of Health Sciences*, 3(2), 249-258.
- Stukalina, Y. (2012). Addressing service quality issues in higher education: the educational environment evaluation from the students' perspective. *Technological and Economic Development of Economy*, 18(1), 84-98
- Talukdar, J., Aspland, T., & Datta, P. (2013). Australian higher education and the Course Experience Questionnaire: insights, implications and recommendations. *Australian Universities Review*, 55(1), 27-35.
- Tucker, B., Jones, S., & Straker, L. (2008). Online student evaluation improves Course Experience Questionnaire results in a physiotherapy program. *Higher Education Research & Development*, 27(3), 281–296.
- Vermunt, J., & Magidson, J. (2005). Technical guide for Latent GOLD 4.0 Basic and Advanced. Belmont, MA: Statistical Innovations Inc.
- Wedel, M., & Kamakura, W. (2000). Market Segmentation: Conceptual and Methodological Foundations (2nd ed.). Boston, MA: Kluwer Academic.