

# **Improved student satisfaction with teaching quality in VET: how it may have happened**

*Peter Murphy, RMIT University*

## **Abstract**

Using the Good Teaching Scale of the NCVER Student Outcomes Survey for the years 2004 and 2005, two TAFE programs at a Victorian dual-sector university were identified as having an increase of approximately 20% or more in student satisfaction ratings with teacher performance in 2003-2004. To investigate what had happened in these programs to produce such increases, semi-structured interviews were conducted with staff who taught and/or managed them. The research showed that while there was no one key causal factor which explained the increases, staff suggested several factors as possible explanations for the increase: a personnel change in the key role of program manager; the introduction of new teaching staff, which had both a positive and a negative effect on the GTS scores; and differences in the student cohort itself. No specific efforts were made by staff to increase the GTS score in 2003-04.

## **Introduction**

Feedback from clients or customers is considered important by many organisations and Australian universities and TAFE institutes are no different, with students considered as their primary 'customers'. Student satisfaction surveys are a common instrument for collecting feedback from students on a range of issues and concerns and have been around for several decades. In Australia, in the Technical and Further Education (TAFE) sector, the common benchmark survey is the Student Outcomes Survey (SOS). The SOS is managed by the National Centre for Vocational Educational Research (NCVER) and is conducted every two years. Within the SOS, the first 6 question items are about teaching quality and performance and are known collectively as the Good Teaching Scale (GTS). While a rough, simple measure by any standards, the GTS is the only national benchmark on teaching quality available within the TAFE system. Therefore it is a significant measure for anyone concerned about teaching quality in TAFE.

It is self-evident that any teaching organisation is interested in improving teaching quality. It may also be assumed that teaching organisations take seriously the feedback received from students about the quality of teaching. Therefore, when student surveys show a marked improvement in the level of satisfaction students have with the teaching that they have received, this is of interest and begs the question "How did the improvement come about?" The current study seeks to address this question.

## **Literature Review**

Student evaluations of teaching (SET) go by many names (eg teacher rating forms, student ratings of instruction, teacher course evaluations, etc). No matter what name they go by, they have been part of academic life for many years (Aleamoni, 1999; Algozzine, Beattie, Bray, Flowers, Gretes & Howley, 2004).

Consequently there is an extensive body of research on SETs and their use in the tertiary education sector. Unfortunately the vast majority of this research is based in the university sector, with little done in VET. Mitchell et al (2006), for example, in a major report on quality as a critical issue in teaching and learning in VET, makes no mention of SETs. One of the few pieces of research that does use SETs found strong student endorsement of the quality of teaching in VET but did not investigate any improvements over time nor how they might have happened (Polesel, Davies, & Teese, 2004). Given the paucity of VET research on SETs, this literature review will hereafter consider research done on SETs in higher education. This is done on the assumption that many of the issues about SETs would be common between the two sectors. However, it must be acknowledged that this assumption is open to question.

The higher education research on SETs is a body of research extensive enough for some to claim that SETs have been “studied to death” (Cohen, 1990; p.123). Not surprising given that even as long as 20 years ago, Cashin (1988) counted over 1800 different studies on the topic and the pace has not let up since then (for a sample of more recent research in the field see Abrami, Theall, & Mets, 2001; Aleamoni, 1999; Algozzine, Beattie, Bray, Flowers, Gretes, Howley, et al., 2004; Centra, 2003; Davidovitch & Soen, 2006; Emery, Kramer & Tian, 2003; Hendry & Dean, 2002; Hounsell, 2003; Kember, Leung & Kwan, 2002; Knapper & Cranton, 2001; Lewis, (ed), 2001; Moore & Kuol, 2005; Ngware & Ndirangu, 2005; Penny, 2003; Richardson, 2005; Ryan, (ed), 2000; Shevlin, Banyard, Davies and Griffiths, 2004; Spooren & Mortelmans, 2006; Yao & Grady, 2005; Zabaleta, 2007).

The nature of this research can be broadly grouped into several themes: the nature of the instruments used in SETs (Richardson, 2005; Sproule, 2000; Theall & Franklin, 1990) how valid they are as a measure of teaching effectiveness (Abrami, 2001; Adams, 1997; Aleamoni, 1999; Kulik, 2001; McKeachie, 1997; Ory & Ryan, 2001; Shevlin et al 2004); potential or perceived bias in their use (Cashin, 1990; Theall & Franklin, 2001); how useful they are for both formative and summative purposes (Davidovitch & Soen, 2006; Kember, Leung & Kwan, 2002; Moore & Kuol, 2005; Murray, 1997; Narasimhan, 2001; Spooren & Mortelmans, 2006; Theall & Franklin, 1991; Theall & Franklin, 2000) and various reviews of the literature or collections of papers on SETs (Abrami, Theall & Mets, 2001; Aleamoni, 1997; Algozzine et al, 2004; Cashin, 1988; Knapper & Cranton, 2001; Marsh, 1987; Richardson, 2005; Ryan, K.E., 2000; Wachtel, 1998).

There continues to be robust debate and discussion about the findings of this extensive body of research and what conclusions can be drawn about SETs and their use. Algozzine et al (2004) state that “Student evaluation of teaching is a very complex and controversial issue with inconsistent research findings” (p.138) while Kulik (2001) states that some studies on SETs are “conflicting, confusing, and inconclusive” (p.10). Nevertheless, Kulik concludes by agreeing with other studies that claim to show that SETs are reliable, can be validly used as a measure of teaching effectiveness and are useful in improving teaching (Centra, 2003; Marsh, 1987; Penny, 2003; Spoorens and Martelman, 2006).

However, of all the studies considered only one directly addressed the issue of how SET scores might be improved. Neath (1996), somewhat tongue-in-cheek in places, lists twenty 'tips' whereby someone could increase their SET ratings without having to 'bother' with improving their teaching. These include such items as being male, grading leniently, being entertaining, being like your students in terms of attitudes, beliefs and values and picking successful students to teach. Tongue-in-cheek or not, many of Neath's tips are seriously challenged by Aleamoni (1999). Through a review of the research, Aleamoni critically assesses 16 myths about SETs, myths which cover similar territory to Neath's 20 tips. He concludes that, on the whole, the myths are just that, myths.

## **Research Method**

This mixed method study (Creswell, 2003) was undertaken in three stages. The first stage was quantitative and involved identifying TAFE programs where there had been a marked improvement in the GTS over a two-year period (the term 'program' is used, rather than the more commonly used 'course', as this is the nomenclature used at RMIT University, the location of the study). RMIT University is a dual-sector university (i.e. it offers both TAFE and higher degree qualifications) in Melbourne, Australia. It was chosen for the study as SOS data is available for every year at RMIT, not just every second year; this means that GTS scores for consecutive years can be compared.

To enable a comparison between programs a simple measure was constructed as follows. Items in the SOS use a standard 5-point Likert scale, with responses ranging from 'strongly agree' to 'strongly disagree'. If the number of 'agree' and 'strongly agree' responses are added together, then divided by the total number of responses overall, a figure is arrived at for each program. Converting to a percentage gives a measure of the GTS. For example, if five students in a particular course complete the survey then there are 30 maximum possible responses to the GTS (5 students x 6 GTS items). If the total number of 'agree' or 'agree strongly' responses are 15 then the GTS score for this course would be  $15/30 \times 100 = 50\%$ .

Using such a GTS measure, two programs were identified as having achieved an increase of approximately 20% or more in their GTS score during the academic teaching period 2003-04. Program A, which had a GTS increase of 18.9%, was a diploma in a visual arts discipline. Program B, which had a GTS increase of 31%, was a Certificate IV in printing. Approaches to staff in both programs saw six Program A staff agree to participate in the study, out of approximately 8 equivalent full-time staff. Two of these staff were in management roles while four were teachers, with the managers both being female while the teachers were all male. In Program B, two staff became involved out of approximately 9 equivalent full-time staff. One was a manager and the other a teacher, both being male. Given the number of staff participating from each program this study is best considered as primarily a study of Program A, with Program B providing supplementary data (Table 1 gives some basic data on the two programs). All staff participants were 50 or older and had extensive teaching experience (from 10-30 years within the TAFE system).

**Table 1 Basic data about Program A and Program B**

	For 2003 academic year	For 2004 academic year
Program A (diploma level)	GTS score 67.8%  No. of survey responses = 20  No. of students completing program = 38 which was 21% of all enrolled students	86.7%  14  No. of students completing program = 25 which was 15% of all enrolled students
Program B (certificate IV level)	GTS score 56.7%  No. of survey responses = 10  No. of students completing program = 19 which was 10% of all enrolled students	87.7%  7  No. of students completing program = 23 which was 32% of all enrolled students

The data in this table was drawn from the Student Outcomes Surveys for 2004-05 and information provided by the Statistics and Reporting Unit of RMIT University.

The second stage of the study was qualitative and involved collecting data from the perspective of the teaching and managerial staff. This was done through semi-structured interviews with all eight participants. All interviews were conducted using a similar format. Once basic demographic data had been obtained (eg sex, age, years of teaching experience), interviewees were told about the gains in the GTS scores over the two-year period under consideration and asked how, from their perspective, the GTS improvements had been achieved. Interviews lasted from 40-50 minutes and were held in a variety of locations (eg the participant's office, the researcher's office, etc).

Once interviews had been transcribed, transcripts were given to interviewees for their consideration and checking. A second interview then took place with each participant where they were given the opportunity to change or alter the transcript of their interview. Apart from a few factual or typographical errors, no changes were made to the original transcripts. Participants were also asked the following additional questions: first, for Program A, to comment upon/confirm the chronology of events over the 2003-04 period (as there were differences of opinion about the exact timing of certain key events); second, to comment on some of the main themes that had arisen from the first round of interviews.

The third stage of the study, also qualitative, involved analysis of the interview data. This began while data was still being collected (Bogdan & Biklen, 1998; Merriam, 1998). This allowed for topics and issues emerging from earlier interviews to be used to inform questions asked at later interviews, bringing a greater focus on key issues to each subsequent interview. Categories of data were identified using open coding, a technique where a term is used to summarise each line of interview data. When all the data has been summarised in this way, the summaries are then reviewed and examined in order to discover commonalities within the data. This was done first within each participant's

own interview data and then broadened in scope to include a search for commonalities amongst all the study's participants. In this way, themes in the data are revealed (Patton, 2002; Strauss & Corbin, 1998). These themes then form the basis for the findings of the study.

## **Findings and Discussion**

With regards to the key research question of how the improved GTS scores happened, this study makes three findings. First, there was no one clear agreed upon factor that explained the improved GTS scores. Second, several factors were suggested by staff participants as possible reasons for the increased GTS score: a personnel change in the key role of program manager; the introduction of different teaching staff; and differences in the student cohort itself. Third, most staff participants stated that no specific conscious effort was made to improve GTS ratings. These three findings will now be discussed in detail.

### **“I don't think any particular thing stands out”**

When first asked how the large increase in the GTS score from 2003 to 2004 had occurred, staff participants gave a range of responses. Some made statements such as “No idea”, “I don't know” and “I can't explain the reason for the jump (in GTS scores)”, indicating a sincere puzzlement over the GTS increase. At first I thought that maybe memory was at fault here, given that I was asking them to recall events 3-4 years in the past. Therefore I pressed them to try and recall 2003 and 2004, asking them for more specific detail about the differences between the two years. While some came up with several further suggestions about what might have happened (all along the lines of what follows in the discussion below), all who had initially said “I don't know” emphasised that these were tentative suggestions and concluded their interview by reiterating that the increase in the GTS score was something of a mystery to them. However, other participants were more certain in their recollections and opinions of what had happened over the two years in question and had no trouble identifying factors they believed strongly contributed to or caused the GTS increase (see below). Therefore, given such varied responses to the key question of how the improved GTS scores were achieved, this study finds that no one key factor can be identified as being responsible for them.

### **“We had a change of coordinator (to someone who was) ... much more capable, much more equitable, much more efficient”**

The position of program manager (PM) is an important leadership role and can have a large influence in how well (or badly) a program runs. Program A suffered the loss of its PM in somewhat dramatic circumstances i.e. at the beginning of second semester 2003 the PM left suddenly due to a serious illness, not returning until the beginning of 2004. In addition, another experienced teacher left at the same time, also due to ill health, and was absent for the rest of 2003. Anybody suddenly called upon to take over the role of PM in such dramatic circumstances would find themselves in a difficult and challenging situation. In many respects, 2003 was the program's 'annus horribilis'.

In contrast, in 2004, once the new PM had implemented changes to the management of the program, described in a positive way as making the program “tighter, more efficient”, then this could be seen as having a positive effect on the 2004 GTS. This could be the case especially because the “tighter, more efficient” comment related in particular to the handling of student issues. For example, marked assessment tasks were returned more promptly. This “helped give students better feedback, a better sense of how they were evolving as a student in the program”. ‘Feedback to students’ is an item in many SET questionnaires (Richardson, 2005) so any perceived improvements in giving students feedback are likely to register positively in the GTS score.

**“We got strangers in” / “we got new young staff in”**

There were two ways, one positive and one negative, in which a change in teaching staff could have affected the GTS score. The negative impact could have come about due to the sudden loss of two teaching staff that happened half way through 2003. It was described variously as “(the loss of the two staff) threw the program into turmoil ... students don’t adapt well to change”, and “... (the students) reacted quite badly because we had to replace their staff with sessionals” and “ ... it was all over the place” and “ ... we all felt that last six months of 03 was awful ... the staff were ... feeling very shaky in 03”. This is strong and emotive language. It indicates the strength of the impact the loss of the two staff had on the team and the program half way through 2003. Therefore, it can be seen how this could have negatively influenced the GTS for that year.

Exacerbating the situation was the use of sessional staff who were brought in to replace the two teachers missing due to illness. Sessional staff per se are not the problem; however, in this instance, due to the work being of a specialist nature, it was difficult finding suitably qualified substitute teachers and so “ ... some of these guys (the replacement teachers) had never taught before.” Consequently, “... the kids really suffered ...” and “... we had a lot of students leave (in 2003)”. This would have been reflected in the low GTS score for 2003.

With regards to the positive impact staff changes had, new staff continued teaching in the program in 2004 but were given professional development to improve their teaching skills. These were subsequently viewed as young staff “... bringing new ideas ... and industry contacts” and “... bringing (a) refreshed focus on industry”. This was cited as having a positive effect on the whole teaching team, reflected in the improved GTS score for the 2004 year.

**“(we had) ... a group of phenomenally clever and focussed students”**

It is a given that student cohorts change from year to year. Their performance is affected by many different variables – gender, educational background, cultural background, age, motivation, literacy skills – to name but a few (Theall & Franklin, 2000). While over a five-year period or longer one might expect some statistical averaging of the impact of such factors, within a short two-year period it is possible that there might be a large difference in the student cohort from one year to the next. When referring to the 2004 academic year, staff used phrases such as “I can think of some wonderful students ... who were there at that time” and “... we’re getting them in at a higher skills level. They succeed

and they're happy about it." Coupled with this comment about 2003, "We had two staff go off very sick ... threw the program into turmoil ... students don't adapt well to change", then a picture emerges of not simply two different student cohorts in terms of their capabilities but two different learning environments and experiences for the two cohorts.

One of Neath's (1996) 20 tips for improving student evaluations is to ensure that you have bright, successful students in your classes. The thinking here is that students who, for whatever reason, find the program difficult and perform poorly tend to project poor performance onto the teaching staff ('if I can't understand this it's because of your poor teaching, not my lack of ability'). Alternatively, if a particular cohort has greater discipline, motivation and maturity, or even simply better natural skills and abilities, and thus performs better, then this is reflected in improved GTS ratings.

**“ no, nothing was done (in response to the poor 2003 GTS score)”**

The final finding of this study was that no conscious effort was made to improve GTS scores over the two year period, other than incidentally through the continuous improvement measures normally undertaken each year. Nearly all participants in the study answered no when questioned about this. I was at first surprised by this, thinking that the increase in 2004 must have been a result of specific measures undertaken by the staff to address the poor 2003 score.

However, upon reflection this could not have been possible for two reasons. Firstly, the results of the SOS are not usually available until the second half of each year. Therefore the relatively 'poor' result of 2003 would not have been known by program teams until well into the second half of 2004, too late to make changes to a program in an attempt to improve these ratings. Secondly, the survey of the 2004 academic year was markedly different to the preceding one in its design. Not being able to validly compare 2004 results with 2003 meant that there was no good reason to suspect that 2003 was a particularly low score but rather that the 2004 score should be considered as a new benchmark (as a result of the new survey design).

**Limitations**

This study had several limitations. Firstly, given that it was a small study covering only two TAFE courses it cannot be generalised to the broader tertiary education sector, other than perhaps in Eisner's sense of generalisation where "Generalizations in education ... need to be treated as tentative guides, as ideas to be considered, not as prescriptions to follow." (1998, p.209). Secondly, several possible ways to triangulate the data were not used due to a number of reasons. Due to a lack of time and resources it was not feasible to interview students who had attended either Program A nor Program B. Similarly, it was not possible to compare GTS ratings with how teaching performance was rated in RMIT's own internal student survey instruments. Staff-student consultative committee minutes would have provided interesting comparative data; however, these were not available for either program. Finally, it would have been informative to analyse the student cohorts of the two programs under study but, again, there was neither the time nor resources to do this.

## **Conclusions**

This study began with a simple question – how did the marked increase in GTS scores over the 2003-04 period in two TAFE programs come about? Through interviewing managerial and teaching staff, and through an examination of the extensive literature on SETs, several conclusions can be drawn. First, like many things to do with teaching, student satisfaction is a complex phenomenon, with multiple variables affecting how students rate their teachers. Measures of student satisfaction are an important indicator of teaching quality within VET but cannot be seen as the only one; cautious use should be made of them. Second, it was surprising to find so few research studies done on SETs and their use in the VET sector. Undoubtedly student feedback on teaching is valued and used, in some way, at the individual institute level. The lack of research about exactly how it is used is a matter for concern. Finally, it seems likely that a combination of factors, all of which played a part, lead to the observed increase in GTS scores from 2003 to 2004, with 2003 being a bad year for Program A in particular. This conclusion is supported by subsequent GTS scores for Program A, all of which are similar to the 2004 result.

In addition, there are two important lessons for other VET practitioners that can be drawn from this study. First is the importance of succession planning for senior management roles. While the sudden unexpected loss of a senior manager might test any organization, no one was qualified and/or experienced enough to replace Program A's Program Manager without it having a substantial negative impact on the program and its students in 2003. Such an impact could have been ameliorated had other staff working in the program been readied and prepared for a more senior role. Second is the importance of teaching qualifications. It is not enough that VET staff be excellent in their chosen profession or trade, they also need to be good teachers so that students can benefit from their expertise. While some people are naturally gifted in this regard this cannot be assumed, and teaching qualifications, while not necessarily guaranteeing a good teacher, do at least indicate that someone has a grasp of the basics of how to teach. Neither of these points are particularly novel but rather reminders of how important it is to get the basics of good teaching and good management right.

## **Acknowledgements**

I gratefully acknowledge the award of a 2007 Learning and Teaching Investment Fund Teaching Fellowship by RMIT University. The fellowship gave me the opportunity to conduct this study. I'd also like to thank five other RMIT staff involved with the teaching fellowships who gave me invaluable support and feedback at various stages of the project: Karen Nankervis, Josephine Lang, Madeleine Shanahan, Wendy Forrest and Sandra Jones. Finally, I'd like to thank the eight RMIT staff members who participated in this study for contributing their time and thoughts.



## References

- Abrami, P.C. (2001). Improving judgments about teaching effectiveness using teacher rating forms. *New Directions for Institutional Research*, 2001:109, 59-87.
- Abrami, P.C., Theall, M. & Mets, L.A. (eds) (2001). *New Directions for Institutional Research*, Spring 2001:109.
- Adams, J.V. (1997). Student evaluations: The ratings game. *Inquiry*, 1:2, 10-16.
- Aleamoni, L.M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*. 13:2, 153-166.
- Algozzine, B., Beattie, J., Bray, M., Flowers, C., Gretes, J., Howley, L., et al. (2004). Student evaluation of college teaching. *College Teaching* 52:4, 134-141.
- Bogdan, R. C., & Biklen, S. K. (1998). *Qualitative research in education: An introduction to theory and methods*. Allyn & Bacon. Needham Heights, MA.
- Cashin, W. E. (1988) *Student ratings of teaching: A summary of research*. IDEA Report No. 20, Kansas State University, Division of Continuing Education.
- Cashin, W.E. (1990). Students do rate different academic fields differently. In Theall, M. & Franklin, J. (eds) *Student Ratings of Instruction: Issues for Improving Practice*. New Directions for Teaching and Learning. No 43. Jossey-Bass, San Francisco.
- Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, 44:5, 495-518.
- Cohen, P. (1990). Bringing research into practice. In Theall, M. & Franklin, J. (eds) *Student Ratings of Instruction: Issues for Improving Practice*. New Directions For Teaching And Learning, Fall 1990:43, 123-132.
- Creswell, J.W. (2003). *Research Design. Qualitative, Quantitative and Mixed Methods Approaches* (2nd Edition), Sage Publications, Thousand Oaks.
- Davidovitch, N. and Soen, D. (2006). Using students' assessments to improve instructors' quality of teaching. *Journal of Further and Higher Education*, 30:4, 351-376.
- Eisner, E.W. (1998). *The Enlightened Eye: Qualitative Inquiry and the Enhancement of Educational Practice*. Prentice-Hall, New Jersey.
- Emery, C.R., Kramer, T.R. & Tian, R.G. (2003). Return to academic standards: a critique of student evaluations of teaching effectiveness. *Quality Assurance in Education*, 11:1, 37-46.
- Hendry, G.D. & Dean, S.J. (2002). Accountability, evaluation of teaching and expertise in higher education. *International Journal for Academic Development*, 7:1, 75-82.
- Hounsell, D. (2003). The evaluation of teaching. In Fry, H., Ketteridge, S. & Marshall, S. (eds). *A Handbook for Teaching & Learning in Higher Education: Enhancing Academic Practice* (2<sup>nd</sup> ed). RoutledgeFalmer, London.
- Kember, D., Leung, D.Y.P. & Kwan, K.P. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment & Evaluation in Higher Education* 27:5, 411-425.
- Knapper, C. & Cranton, P. (eds) (2001). Fresh approaches to the evaluation of teaching. *New Directions For Teaching And Learning*, Winter 2001:88.

- Kulik, J.A. (2001). Student ratings: validity, utility, and controversy. *New Directions for Institutional Research*, 2001:109, 9-25.
- Lewis, K.G. (ed) (2001). Techniques and strategies for interpreting students evaluations. *New Directions For Teaching And Learning*, Fall 2001:87.
- Marsh, H. W. (1987). Students' evaluations of university teaching: research findings, methodological issues, and directions for further research, *International Journal of Educational Research*, 11:3, 253–388.
- McKeachie, W. (1997) Student ratings: The validity of use, *American Psychologist*, 52 (11), 1218–1225.
- Merriam, S. B. (1998). *Qualitative research and case study applications*. Jossey-Bass, San Francisco.
- Moore, S and Kuol, N. (2005). Students evaluating teachers: exploring the importance of faculty reaction to feedback on teaching. *Teaching in Higher Education*. 10:1, 57-73.
- Mitchell, J., Chappell, C., Bateman, A & Roy, S. (2006). Quality is the key: Critical issues in teaching, learning and assessment in vocational education and training. NCVET, Adelaide.
- Murray, H.G. (1997). Does evaluation of teaching lead to improvement of teaching?, *International Journal for Academic Development*, 2:1, 8-23.
- Narasimhan, K. (2001). Improving the climate of teaching sessions: The use of evaluations by students and instructors. *Quality in Higher Education*, 7:3, 179-190.
- Neath, I. (1996). How to improve your teaching evaluations without improving your teaching. *Psychological Reports* 78. 1363-1372.
- Ngware, M.W. & Ndirangu, M. (2005). An improvement in instructional quality: Can evaluation of teaching effectiveness make a difference? *Quality Assurance in Education*. 13:3, 183-201.
- Ory, J.C. and Ryan, K. (2001). How do student ratings measure up to a new validity framework? *New Directions for Institutional Research*, 2001:109, 22-44.
- Patton, M. (2002). *Qualitative research and evaluation methods*, Sage, Thousand Oaks, CA.
- Penny, A. R. (2003) Changing the agenda for research into students' views about university teaching: Four shortcomings of SRT research, *Teaching in Higher Education*, 8:3, 399–411.
- Polesel, J., Davies, M. & Teese, R. (2004). Course completion and instructional experience in TAFE. NCVET, Adelaide.
- Richardson, J.T.E. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment and Evaluation in Higher Education*, 30:4, 143-157.
- Ryan, K.E. (ed) (2000). Evaluating teaching in higher education: A vision for the future. *New Directions For Teaching And Learning*, Fall 2000:83.
- Shevlin, M., Banyard, P. Davies, M. and Griffiths, M. (2004) The validity of student evaluation of teaching in higher education. In Tight, M (ed) *The RoutledgeFalmer Reader in Higher Education*, RoutledgeFalmer, London.
- Spooren, P. & Mortelmans, D. (2006). Teacher professionalism and student evaluation of teaching: Will better teachers receive higher ratings and will better students give higher ratings? *Educational Studies*, 32:2, 201-214.
- Sproule, R. (2000). Student evaluation of teaching: A methodological critique of conventional practices. *Education Policy Analysis Archives*, 8:50. Available at: <http://epaa.asu.edu/epaa/v8n50.html>

- Strauss, A. L., & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage, Newbury Park, CA.
- Theall, M. and Franklin, J. (eds) (1990). *Student Ratings of Instruction: Issues for Improving Practice*, Jossey-Bass Inc, San Francisco.
- Theall, M. and Franklin, J. (eds) (1991). Effective practices for improving teaching. *New Directions For Teaching And Learning*, Winter 1991:48.
- Theall, M. and Franklin, J. (2000). Creating responsive student ratings systems to improve evaluation practice. In Ryan, K.E. (ed) (2000). *Evaluating Teaching in Higher Education: A Vision for the Future*. *New Directions For Teaching And Learning*, Fall 2000:83, 93-108.
- Theall, M. and Franklin, J. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction? *New Directions for Institutional Research*, 2001:109, 45-56.
- Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment and Evaluation in Higher Education*, 23:2, 191-211.
- Yao, Y. & Grady, M.L. (2005). How do faculty make formative use of student evaluation feedback?: A multiple case study. *Journal of Personnel Evaluation in Education*, 2005:18, 107-126.
- Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education*. 12:1, 55-76.