



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

School of Education

Te Kura Toi Tangata

Waikato Journal of Education

Te Hautaka Mātauranga o Waikato

Volume 11 No.2 : 2005



WAIKATO JOURNAL OF EDUCATION

TE HAUTAKA MĀTAURANGA O WAIKATO

Editors:

Deborah Fraser
Toni Bruce

Editorial Committee:

Miles Barker
Margaret Carr
Pat Day
Rosemary de Luca
Alan Hall
Clive McGee
Judy Moreland
Clive Pope

Waikato Journal of Education is a refereed journal, published annually, based in the School of Education, University of Waikato, Hamilton, New Zealand. It publishes articles in the broad field of education. For further information visit the WJE website <http://www.soe.waikato.ac.nz/wje/>

Correspondence should be addressed to: Rosemary de Luca and Toni Bruce, Editors, School of Education, Private Bag 3105, The University of Waikato, Hamilton, New Zealand. Email: deluca@waikato.ac.nz and/or tbruce@waikato.ac.nz

Books for review should be sent to the Editors

Business correspondence: Orders, subscription payments and other enquiries should be sent to the Business Manager, *Waikato Journal of Education*, Wilf Malcolm Institute of Educational Research, School of Education, The University of Waikato, Private Bag 3105, Hamilton, New Zealand.

Subscriptions: Within NZ \$30; Overseas NZ \$40

Copyright: © School of Education, The University of Waikato

Publisher: School of Education, The University of Waikato

Cover design: Donn Ratana

Printed by: Waikato Print

Layout: Gillian Joe and Carolyn Jones

WJE 2006: Call for papers: Pacific education, Research and practice

The Waikato Journal of Education is a well-established peer reviewed publication that has quality articles on a range of topics related to education.

New Zealand has a strong presence in Pacific education, and Pacific communities have a strong presence in New Zealand schools. However, opportunities for publication of Pacific research in mainstream journals are limited. Therefore, this call for papers seeks articles that focus on Pacific education; both research and practice. Pacific research is reflective of the traditions of the past, as well as the present and future. It often embodies different paradigms, perspectives and critical stances that are not always captured in mainstream research and aims to benefit Pacific communities. Articles will be welcomed that theorise about Pacific research, report on research projects, report on an innovative practice or initiative, or a combination of any of these. As well as traditional manuscripts, the journal welcomes submissions in other formats, such as short stories, poetry and drawings.

Submissions please to Timote Varioleti (varioleti@waikato.ac.nz) and Jane Strachan (jane@waikato.ac.nz), School of Education, The University of Waikato, PB 3105, Hamilton. Please submit 3 blind copies and a separate page with author/s contact details by 30 April 2006. Electronic submissions also accepted for consideration.

ISSN 1173-6135

WAIKATO JOURNAL OF EDUCATION

TE HAUTAKA MĀTAURANGA O WAIKATO

VOLUME 11, No. 2, 2005

Will Scholars Trump Teachers In New Zealand Teacher Education?	3
NOELINE ALCORN	
Changing School Culture Through Action Research And Leadership	17
NANCY HIGGINS	
Hauora And Physical Education In New Zealand: Perspectives Of Māori And Pasifika Students	37
KATIE FITZPATRICK	
The Efficacy Of Using A Feedback Typology As A Heuristic Device To Deconstruct Teachers' Feedback Practice	49
HELEN DIXON	
Principles Of Effective Literacy Practice For EAL Students In New Zealand Classrooms¹	67
MARGARET FRANKEN	
Academic Dishonesty Among Students In Tertiary Institutions: A Literature Review	83
KELLY DE LAMBERT, NICKY ELLEN AND LOUISE TAYLOR	
Who Is Learning What From Student Evaluations Of Teaching?	101
LANCE KING	
Talking Across The Divide: English Teachers Respond To The NCEA	113
TERRY LOCKE	
School Quasi-Markets: Best Understood As A Class Strategy?	137
MARTIN THRUPP	

WHO IS LEARNING WHAT FROM STUDENT EVALUATIONS OF TEACHING?

LANCE KING

Education consultant and University of Waikato Masters student

DEBORAH FRASER

*Department of Human Development & Counselling
The University of Waikato*

ABSTRACT *Student evaluations of teaching (or SET) through anonymous survey forms are a consistent practice in higher education across the world yet research results vary considerably as to the reliability, validity and efficacy of SET. Nonetheless, the widespread use of SET for promotion and tenure decisions ensures that these results are high stakes for tertiary staff. The tension between the purposes of SET (to supposedly improve teaching) and the ramifications of SET results are explored. Staff and students tend to hold very different views of SET and the issue of maintaining high academic standards can be at risk. However, SET can be used as an opportunity for staff and students to work together on issues in teaching and learning that enhance quality for all concerned.*

KEYWORDS

Student evaluation of teaching, Tertiary teaching, Teaching quality

INTRODUCTION

Student evaluations of their courses and teachers are now used in almost all colleges, but it is not at all clear that they have had a positive effect. There is much evidence that they have contributed to a decline in the level of intellectual intensity in the classroom. (Trout, 1997, p. 3)

While student evaluations of teaching (SET) were originally intended for formative and diagnostic use, they are most commonly used in a summative fashion. SET data has become increasingly 'high stakes' as the outcomes can have a significant influence on salary and tenure decisions for teaching staff in many universities and colleges. However, SET data is notoriously unreliable and inconsistent and the SET process often challenging and unsatisfying for staff and students alike. Some of the major issues are explored here with reference to research findings on SET survey forms that are commonly used in tertiary education.

THE ORIGIN AND INFLUENCE OF STUDENT EVALUATIONS OF TEACHING

Engstrom (1999) cites 1926 as the birth-date of SET. The birth-place was Washington University where a staff researcher set in motion a sea-change in

thinking in education by administering to those students present *The Purdue Rating Scale of Instruction* (VanArsdale & Hammons, 1995). For possibly the first time ever, the opinions of tertiary students were genuinely sought as to the nature of the teaching they experienced and those students were given the opportunity to have some influence over that same teaching process. This student survey represented a shift in operational ethic at Washington University, from the *benevolent autocracy* ethic of traditional Victorian education to the *individual's right to informed consent* ethic of the American flapper era. SET heralded a revolution in thinking in tertiary education and represented the first true step towards systematising androgogy – designing education for adults by adults.

By the 1980s, in higher education institutions across the world, SET through end-of-course surveys had become the primary means of evaluating teacher performance. As of 1988, 80 percent of all liberal colleges in the USA used systematic student ratings as all or part of the means for evaluating teachers (Langbein, 1994). Across tertiary education in the UK, data from SET is considered as important evaluative information on which to judge academic staff and as a guide for potential changes in course material and method of delivery (Shevlin, 2000). Moreover, in Australian universities it is generally considered that student evaluations are more useful, accurate and valid than other measures of teaching performance and have the added benefit of being a direct measure of ‘consumer’ satisfaction (Ramsden, 1991).

SET is an instrument designed to assess the quality of teaching as experienced by the learner. Arguably, the process of evaluative feedback by students is enhanced by the anonymity of the process where the teacher cannot identify any individual student and the results are confidential to the teacher concerned. Ideally, students are empowered with the ability to have some influence over the teaching process and teachers become empowered with the ability to continually improve the effectiveness of their teaching which improves the effectiveness of the students’ learning, creating a responsive and continuous learning-teaching cycle. In some instances, however, staff development can seem the result of external surveillance when teaching results are monitored and specific teachers are targeted for staff improvement (Black, Cannon & Hicks, 2000).

Problems can arise if anonymity is not maintained and the diagnostic potential can be over-shadowed by administrators and leaders in positions of power requiring the results of teachers’ SET for staff promotion or tenure. While these instances can enhance teacher quality, they take the results out of the realm of personal professional development and introduce high-stakes accountability. The consequences are inevitable:

In the past decade SETs have changed from benign tools by which individual faculty members attempted to improve teaching skills into a mandatory process on which academic administrators rely (sometimes exclusively) to measure teaching effectiveness for salary and promotion decisions. (Stratton, Myers & King, 1994, p. 1)

The use of SET to support curriculum and personnel decisions in higher education as measured across 600 liberal-arts colleges in the US increased from:

- 28 percent in 1973, to
- 68 percent in 1984, and then to
- 86 percent in 1993 (Seldin, 1993).

The direct influence of SET on salary change for academic staff also rose as a proportion of overall influence during that same period as illustrated by the University of Akron which, in 1986, mandated that from that time onwards 45 percent of salary adjustment funds be allocated by teaching effectiveness criteria (Stratton, et al., 1994).

And although nearly 75 percent of academics judge student course evaluations as unreliable and imprecise measures of performance, nearly 100 percent of tertiary institutions use them frequently and exclusively (Reckers, 1995). One of the main reasons is that student evaluations of teaching are easy and cheap to administer and they give the administrators some measure of quality and some influence over teacher performance in the classroom. Teaching staff, however, are often skeptical about the merits of SET and are concerned that the significance given to the results of such data collection may be out of proportion with the validity and reliability of the data itself. The New Zealand Vice-Chancellor's committee survey (2003) acknowledged that SET has a vital part to play in the monitoring of teaching quality but added that relying solely on data like SET is too narrow and that such evaluations should be balanced by range of other assessment tools such as self-evaluations, portfolios and peer appraisals.

VALIDATING SET DATA

The research on the validation of SET data is overwhelming in both volume and difference of opinion. To summarise, everything has an influence on SET scores, from the attitude of the students to the personality of the teacher and a myriad of factors in between. The sheer abundance of research in this area highlights the difficulties that exist in defining the teaching/learning interface and the factors that contribute to success therein.

In a study of 2,121 Accounting students in Austrian Commercial Colleges, Greimel-Fuhrmann and Geyer (2003) found that the students' global rating of teachers was affected by the students' attitudes towards evaluating their teachers, as well as by the students' liking for their teacher and their interest in the subject of Accounting. In contrast, a survey in Provincetown USA of 350 undergraduates at college found that the only significant predictor of high scores in student evaluations was the extroversion of the teacher (Radmacher & Martin, 2001).

Stephen J. Ceci, a Cornell professor of human development, undertook an investigation of the factors that might influence SET. Twice he taught an identical course in developmental psychology he had taught for almost 20 years. The one difference in his teaching was that he used a more enthusiastic tone of voice during the second semester. Student ratings on his teaching at the end of the course soared on every measure for the second term (Schroeder, 1997).

That combination of enthusiasm and extroversion which seems to make for good SET is what Mark Shevlin of Ulster University calls the 'Charisma Factor'. In

a survey of 213 UK social science undergraduate students he was able to attribute 69 percent of the variation in 'lecturer ability' to this charisma factor and he argued that charisma was an example of a single central trait which can seriously effect a student's overall evaluation of the lecturer (Shevlin, Banyard, Davies & Griffiths, 2000). While charisma can correlate with teacher effectiveness it does not necessarily follow that a charismatic teacher improves students' learning. Charisma does not feature in a number of large studies on the qualities of expert teachers (see e.g., Hattie, 2002). What is mentioned time and again is the passion teachers have for the subject they teach and the desire to share this knowledge (Brookfield, 1995; Carpenter, McMurchy-Pilkington & Sutherland, 2002; Ramsden, 1992). Moreover, expert teachers ensure challenge, monitor student learning carefully and provide quality feedback (Hattie, 2002). These qualities are not determined by charisma. Expert teachers who are not naturally charismatic (but who are still deeply passionate about their subject) can be rated lower than a colleague who is high on the 'C' factor but low on ability to challenge students or monitor students' progress and provide feedback.

Gender of both teachers and students has the potential to be a big factor in affecting SET but research results are mixed. Some studies have found no (or extremely small) differences between the evaluation of female and male teachers on the basis of student gender alone (Centra & Gaubatz, 2000). Other studies report gender bias, with male students rating female instructors lower than male instructors (Basow, 1995).

The experience of and exposure to the teacher have also been shown to influence SET scores, with a positive correlation between student evaluations and

1. the percentage of time teachers spent in instructional behaviours (demonstrated by Engstrom, 1999, in the Physical Education environment) and
2. the experience of the teacher, with SET scores increasing up to a peak at 13 years of teaching experience and then trending downward again (Langbein, 1994).

The overall effect of these variables, although important, does not invalidate the evaluation process and we must be aware that the criticism of SET is itself not necessarily unbiased – "the search for potential biases to student ratings has itself been so biased, that it could be called a witch hunt" (Marsh, 1987, p. 253).

According to these findings, in order to score high on SET a teacher would need to be charismatic, enthusiastic, extroverted, in the job 10-13 years and prepared to spend significant amounts of time deliberately instructing students. Nothing surprising there. However, the extent to which a high SET score leads to increased student achievement is another issue again.

IMPROVING THE ACHIEVEMENT OF STUDENTS

The issue of student achievement is the cornerstone of the largest area of debate in the SET literature. This literature examines two major questions: Do good SET ratings correlate with good student grades? If so, are these good student grades the result of improved teaching quality or the lowering of grading standards?

Yunker and Yunker (2003) cite a considerable body of research showing a positive correlation between student evaluations of staff members and objective measures of student achievement before evaluating the results of their own research with Accounting students which showed the exact opposite. With their own students the authors found a statistically significant negative relationship between student evaluations and student achievement. These results contradict work by Stapleton and Murkison (2001) who demonstrated a positive relationship between teacher excellence scores and level of learning achievement.

Stapleton and Murkison's findings are supported by Stratton et al. (1994) who, in the Economics department of the University of Akron, demonstrated an initial increase in grades of about 11 percent on the introduction of SET which unfortunately was followed by nine years of downward trending grades – once adjusted for 'student quality'. Of the teachers who had improved students' grades since the introduction of SET it was claimed to be impossible to tell whether that improvement had come about through improving teaching methods or through the lowering of grading standards (Stratton et al., 1994).

The potential for the lowering of grading standards and the concomitant issue of unrealistically raising grade expectation is the focus of research in this area, and two interesting quantitative studies highlight the issues. The first is Laura Langbein's survey of 2600 American university students where she showed that for each additional unit increase in the expected grade (e.g., B to A or C to B), the total scale score (SET) increased by 3.36 points. She also related SETs to actual academic grades achieved and was able to show that for each additional unit increase in actual GPA (e.g., from 3.0 to 4.0) the total scale score decreased by 1.00. So positive expectations were having a positive influence and successful learning was having a negative influence on students' judgement of teaching effectiveness (Langbein, 1994).

The second is Allen Kelley of Duke University's study which showed that the students' expectation of their course grades had a positive and statistically significant impact on both course and professor ratings but only of 2 and 3 percent respectively (Kelley, 1972).

STUDENT AND STAFF VIEWS OF SET

As would be expected, academic staff and students have widely differing views on the validity of the research findings and on SET itself. According to Sojka, Gupta and Deeter-Schmelz (2002), staff believe that students award easier, more entertaining teachers with higher ratings and that students do not take SET seriously. Students strenuously deny this. Forty-five percent of students think that professors do nothing with SET information. Staff would like to reduce the

weighting given to SET whereas students would like to increase it and have SET as central in all promotion, salary and tenure decisions (Sojka et al., 2000). The majority of students in the Sojka et al. study and in a study by Dwinell and Higbee (1993) were not aware of the power and effect of SET scores and did not believe that their evaluations affected salaries or promotion and tenure decisions.

Staff who receive negative feedback from SET are often personally affronted by the results of the process, calling comments made about them 'stinging' and even 'downright rude' (Bodle, 1994), and believe that students punish teachers for using 'well-known' learning/teaching techniques. This, they believe, then encourages teachers to increase SET scores by 'sacrificing' the learning process. Some staff members even go so far as to assert that "the anonymity of the SET process encourages students to attack certain teachers without fear of punishment, and to damage professional reputations without giving the staff members the ability to 'seek restitution'" (Crumbley, Henry & Kratchman, 2001, p. 206).

The personal difficulties teaching staff have with the SET process are exemplified by feedback from those involved in the *Making Sense of Learning at Secondary Schools* project presently undertaken by Massey University in three New Zealand secondary schools through 2004-5. Even though the setting is secondary, rather than tertiary, some of the anxieties and shifts in thinking required are the same. For example, a teacher commented:

... I had to come to grips with (the idea) that something that would be valuable for me would be to hear what my students actually felt, what they thought about the process through which I taught them, the process in which they actually participated in the class. (Kane & Maw, 2005, p. 317)

Indeed, teachers at all levels need to avoid defensiveness in order to learn from student feedback, especially when that feedback may be rather challenging or unpalatable. It can take considerable effort for even experienced teachers to see their teaching through the eyes and minds of their students. Students, however, tend to be very clear about the merits of doing so. For example, a Year 12 student in the same study commented:

Maybe if the teachers or the school people actually really listened to us they could have a deeper, much clearer view and understanding of what we are going through in school. (Kane & Maw, 2005, p. 321)

POTENTIAL, POSSIBILITIES AND PITFALLS

While it is unlikely that there is an ideal SET process for all institutions, it is useful to look at the basis on which one can be constructed. Coralie McCormack of the University of Canberra has produced an ethical framework within which SET can be constructed. Her four principles are:

1. *close the feedback loop with students* – communicate purpose, process, outcomes, responses and past experiences with SET

2. *students should be aware of, and consent to, the use of their feedback for research and/or publication beyond the bounds of the individual subject/course evaluation* – obtain all students' informed consent
3. *ensure student feedback questionnaires are fairly administered* – using standardized procedures to safeguard the validity and reliability of the information obtained and
4. *ensure anonymity of all evaluative data collected* – no student ID required (McCormack, 2005).

Once students are reassured about the integrity of the process, then if SET results are removed from the purview of administrators and made confidential to the teacher/instructor/professor concerned and their professional development counterparts (such as a buddy in support, mentor, tertiary teaching adviser and/or critical friend) then some of the concerns of staff will also be resolved. This does not preclude staff from using SET results in their promotion applications but it puts greater emphasis on teachers' professional judgement and responsibility. It might not be appropriate to submit SET results in some courses in some years due to a wide range of issues from illness of the teacher at some crucial stage during a course through to major changes in course design that require refinement. With the emphasis on ongoing improvement, staff can use results to work on the feedback received. If, however, they know that all results will be viewed by those deciding their career promotion, they could be more concerned about obtaining high SET scores than improving teaching and learning. Moreover, if SET were carried out within rather than at the end of any course, then students would have the opportunity to see direct evidence of whether their feedback was listened to or not. Such an approach is advocated by Brookfield (1995) who employs more informal versions of SET through anonymous *Critical Incident Questionnaires*. These surveys are intended for use during a course of study in order to ascertain students' responses to the teaching and adjust where necessary those teaching practices that require modification in order to maximize learning. Adaptations like this are difficult to institute where SET has been institutionalized as a process of mutual mistrust but there is considerable scope to create a SET system that has learning and teaching benefits for all concerned.

The concerns of staff and students about the necessity, validity and usage of SET data have made it very difficult for some educational institutions to settle on one measurement tool which is acceptable to all parties. By the year 2000 across Auckland University, for example, there were six different SET instruments in use and departments had a seventh option to design their own (Brown & Frielick, 2000). The actual design of data gathering instruments for SET is beyond the scope of this paper but the experience of two Australian universities using different instruments may be of use.

In 1983 at the University of Queensland, TEVAL (Teaching Evaluation) was designed and implemented as the first generic teaching evaluation instrument to be used within Australia. Since its introduction there have been three evaluations of TEVAL conducted by the university, as reported by Timpson and Andrew (1997). The first was by Moses (1988) who found that most academics were in favour of

using TEVAL and 96 out of 104 had used TEVAL results to make positive changes to their teaching practice. The second, by Baxter (1991), found overall satisfaction with TEVAL had dropped to 84 percent with only 44 percent of academic staff who felt TEVAL results had a great or considerable effect, and 47 percent who felt they had a slight effect. By 1992, Ramsden and Martin found the approval of TEVAL had dropped to 63 percent, which they attributed to the increasing use of TEVAL results for personnel purposes. These studies led to concerns that the single, all purpose instrument was not able to distinguish adequately between teaching and subject or course evaluation and, by 1995, TEVAL had been replaced by three different instruments – i) Teaching Feedback; ii) Subject Evaluation; and iii) Approaches to Studying. The separation of evaluation into these three components and a refocusing on the use of teaching portfolios for appraisal purposes enabled Timpson and Andrew (1997) to report “more attention to teaching generally; more support for instructional innovation and improvement; and more acceptance of a multifaceted assessment of teaching effectiveness” (p. 56). Since 1997 there does not appear to have been any further research published on the effectiveness of the new system.

In 1996 the imminent demise of the software package used to support SET caused Dr Duncan Nulty of Queensland University of Technology (QUT) to develop WOLF or (Web On-Line Feedback) which, by 2000, was operational in five Australian tertiary institutions. WOLF was designed to enable any staff member to simply and easily create an on-line feedback questionnaire, to eliminate the feedback delays inherent in a ‘paper and pencil’ system, to improve the range of feedback mechanisms available to staff, to integrate evaluation into the daily practice of teaching professionals and to enable the evaluators to also receive feedback on the action taken as a result of their input. Unfortunately, by 2000 only 166 on-line questionnaires had been created compared with 7865 pen and paper questionnaires, indicating a general lack of uptake of the new technology (Nulty, 2000). The poor response of staff and students to on-line evaluations was also found to be the case at Murdoch University where, even with no other options available, the response rate on-line was only 30 percent. Offering cash prizes moved the response rate up to 54 percent but, not surprisingly, did not prove sustainable. However, a number of useful strategies for creating successful on-line evaluation instruments are offered by Cummings and Ballantyne (2000).

Also at QUT, an example of the use of SET to inform professional staff development was created by Ballantyne, Borthwick and Packer (2000) as a special project implemented by the School of Professional Studies. In this project 87 staff and 127 students were surveyed as to their experience of the teaching/learning dynamic and asked to focus on areas of possible improvement. Although students heavily favoured areas for improvement mostly under staff control and staff favoured areas for improvements largely under student control, six common themes were identified. Collaborative staff-student groups were then formed to address each theme. Each group produced a range of strategies and practical suggestions for overcoming common problems. Booklets were produced by all the groups which were then edited and the series published for use within the University – the series

was called *Enhancing Teaching and Learning* (Ballantyne, Borthwick & Packer, 2000). In this fashion, students and staff aligned themselves together against problems and issues in the teaching-learning nexus. Such a collaborative approach mitigates the effects of SET that do little to improve practice and may put quality at risk. Unfortunately this project appears to have been a 'one-off' and since 2000 follow-up research does not seem to have been undertaken into any on-going evaluative process or the efficacy of the booklets produced. Nonetheless, this example does provide an avenue for further exploration and research in the New Zealand context especially given that collaboration between teachers and students is a strong theme in New Zealand's latest tertiary education policy priorities (Ministry of Education, 2005).

CONCLUSION

While there are many ways in which teachers can seek student evaluation of their teaching, anonymous survey forms used summatively still tend to dominate tertiary education courses. SET are often used for promotion and tenure decisions and are intended to gauge teacher effectiveness but the research indicates a number of issues regarding the validity of such data.

However, as an androgogical tool, SET can become a forum for discussion about questions central to the teaching-learning process. A number of initiatives and suggestions described in the latter part of this paper outline ways in which SET can aid both staff and students in reconsidering, reconfiguring and reevaluating what happens in classrooms to maximize learning and improve teaching. Ongoing research into the short and long term effects of such initiatives will enable an analysis of ways in which teaching quality can be acknowledged and enhanced for the benefit of all concerned.

REFERENCES

- Ballantyne, R., Borthwick, J., & Packer, J. (2000). Beyond student evaluation of teaching: Identifying and addressing academic staff development needs. *Assessment and Evaluation in Higher Education*, 25(3), 221-237.
- Basow, S. A. (1995). Student evaluation of college professors: When gender matters. *Journal of Educational Psychology*, 87(4), 656-665.
- Baxter, E. P. (1991). The TEVAL experience 1983-1988: The impact of student evaluation on university teachers. *Studies in Higher Education*, 16, 151-178.
- Black, B., Cannon, R., & Hicks, O. (2000). *The missing link: Developing the nexus between student feedback surveys and development for teachers*. Proceedings of the HERDSA Conference, Perth, Australia. Retrieved 25 November 2005 from: <http://www.aset.org.au/confs/aset-herdsa2000/procs/black.html>
- Bodle, J. V. (1994). Evaluating student evaluations: The search for perspective. *The Journalism Educator*, 49(1), 76-83.
- Brookfield, S. (1995). *Becoming a critically reflective teacher*. San Francisco: Jossey Bass.

- Brown, J., & Frielick, S. (2000). Student evaluation of teaching at the University of Auckland. In E. Santhanam (Ed.), *Student feedback on teaching: Reflections and projections* (pp. 17-22). Crawley, WA: University of Western Australia Evaluation of Teaching Unit.
- Carpenter, V., McMurchy-Pilkington, C., & Sutherland, S. (2002). Kaiako toa: Highly successful teachers in low decile schools. *SET: Research Information for Teachers, 1*, 4-8.
- Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluation of teaching? *Journal of Higher Education, 71*(1), 17-33.
- Crumbly, L., Henry, B. K., & Kratchman, S. H. (2001). Students' perception of the evaluation of college teaching. *Quality Assurance in Education, 9*(4), 197-208.
- Cummings, R., & Ballantyne, C. (2000). Student feedback on teaching: Online! On target? In E. Santhanam (Ed.), *Student feedback on teaching: Reflections and projections* (pp. 29-37). Crawley, WA: University of Western Australia Evaluation of Teaching Unit.
- Dwinell, P. L., & Higbee, J. L. (1993). Students' perceptions of the value of teaching evaluations. *Perceptual and Motor Skills, 76*, 995-1000.
- Engstrom, D. (1999). Correlations between teacher behaviours and student evaluations in college level physical education activity courses. *Physical Educator, 56*(2), 105-113.
- Greimel-Fuhrmann, B., & Geyer, A. (2003). Students' evaluation of teachers and instructional quality – analysis of relevant factors based on empirical evaluation research. *Assessment and Evaluation in Higher Education, 28*(3), 229-238.
- Hattie, J. (2002, October). What are the attributes of excellent teachers? *Proceedings of the NZCER Annual Conference*, (pp. 3-26). Wellington: NZCER
- Kane, R. G., & Maw, N. (2005). Making sense of learning at secondary school: Involving students to improve teaching practice. *Cambridge Journal of Education, 35*(3), 311-322.
- Kelley, A. C. (1972). Uses and abuses of course evaluations as measures of educational output. *Journal of Economic Education, 4*(1), 13-18.
- Langbein, L. I. (1994). The validity of student evaluations of teaching. *Political Science & Politics, 27*(3), 545-558.
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of International Research, 11*(3), 253-263.
- McCormack, C. (2005). Reconceptualizing student evaluation of teaching: An ethical framework for changing times. *Assessment & Evaluation in Higher Education, 30*(5), 463-476.
- Ministry of Education (2005). *Statement of tertiary education priorities (STEP) 2005-2007*. Retrieved 30 January 2006 from <http://www.minedu.govt.nz/index.cfm?indexID=1028&layout=index>

- Moses, I. (1988). *Academic staff evaluation and development: A university case study*. St Lucia: University of Queensland Press.
- New Zealand Vice-Chancellors' Committee (2003). *Monitoring the quality of teaching*. Wellington: CUAP.
- Nulty, D. (2000). Web on-line feedback (WOLF): Intentions and evaluation. In E. Santhanam (Ed.), *Student feedback on teaching: Reflections and projections* (pp. 38-41). Crawley, WA: University of Western Australia Evaluation of Teaching Unit.
- Radmacher, S. A., & Martin, D. (2001). Identifying significant predictors of student evaluations of staff through hierarchical regression analysis. *Journal of Psychology, 135*(3), 259-268.
- Radner, S. G. (1996). Accountability, evaluation of tenured staff, and program reviews. *Radical Teacher, 48*, 11-16.
- Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in Higher Education, 16*(2), 129-150.
- Ramsden, P. (1992). *Learning to teach in higher education*. London: Routledge.
- Ramsden, P., & Martin, E. (1992). *The student evaluation of teaching service: TEVAL*. Unpublished report to the Tertiary Education Institute.
- Reckers, P. M. J. (1995). Know thy customer. Change in accounting research: A research blueprint. *Journal of Accounting Education, 14*(2), 179-185.
- Schroeder, K. (1997). How students decide. *Education Digest, 63*(4), 73.
- Seldin, P. (1993, 21 July). The use and abuse of student ratings of professors. *Chronicle of Higher Education, 12*, p. 40.
- Shevlin, M., Banyard, P., Davies, M., & Griffiths, M. (2000). The validity of student evaluation of teaching in higher education: Love me, love my lectures? *Assessment and Evaluation in Higher Education, 25*(4), 397-405.
- Sojka, J., Gupta, A. K., & Deeter-Schmelz, D. R. (2002). Student and staff perceptions of student evaluations of teaching: A study of similarities and differences. *College Teaching, 50*(2), 44-48.
- Stapleton, R., & Murkison, G. (2001). Optimizing the fairness of student evaluations: A study of correlations between instructor excellence, study production, learning production, and expected grades. *Journal of Management Education, 25*(3), 269-291.
- Stratton, R. W., Myers, S. C., & King, R. H. (1994). Staff behavior, grades, and student evaluations. *Journal of Economic Education, 25*(1), 5-15.
- Timpson, W. W., & Andrew, D. (1997). Rethinking student evaluations and the improvement of teaching: Instruments for change at the University of Queensland. *Studies in Higher Education, 22*(1), 55-65.
- Trout, P. A. (1997). Student course evaluations: What the numbers mean. *Current, 398*, 3-6.
- VanArsdale, S. K., & Hammons, J. O. (1995). Myths and misconceptions about student ratings of college staff: Separating fact from fiction. *Nursing Outlook, 43*(1), 33-36.

- Yunker, P. J., & Yunker, J. A. (2003). Are student evaluations of teaching valid? Evidence from an analytical business core course. *Journal of Education for Business*, 78(6), 313-317.