

Research challenges inherent in determining improvement in university teaching

Marcia Devlin

Deakin University

Using a recent study that examined the effectiveness of a particular approach to improving individual university teaching as a case study, this paper examines some of the challenges inherent in educational research, particularly research examining the effects of interventions to improve teaching. Aspects of the research design and methodology and of the analysis of results are discussed and recommendations for improvements for future research are made.

In a recent study of the effectiveness of a particular approach to assisting individual university lecturers to improve their teaching (Devlin, 2007), it was necessary to attempt to measure teaching effectiveness in order to determine to what extent the approach used was successful in bringing about the desired improvement. Using this study as a basis for discussion, this current paper examines some of the issues inherent in educational research that seeks to determine the impact and/or effectiveness of interventions to improve teaching and learning. Matters related to the analysis of statistical results, aspects of the research design that may have affected the results and methodological rigour are discussed, among other research-related issues. The implications for those involved in examining the effects of efforts to improve university teaching are outlined.

Designing and conducting research to determine the effects of educational interventions is a challenging endeavour. The study around which the current paper is based represents an attempt to conduct a rigorous empirical research project which incorporated random allocation to intervention and control groups, pre- and post-intervention measures of teaching and learning and the use of psychometrically sound measurement tools. In addition, qualitative data were incorporated into the design to add depth and breadth to the findings.

There is a paucity of rigorous examination of the outcomes of academic development and the study on which this paper is based sought to contribute to the literature in this area. A pretest-posttest control group intervention design with random allocation to group was employed.

The intervention group undertook an individual program of teaching improvement. The control group was used to control for variables other than the intervention that may have contributed to changes to teaching over the one year period of the study. Sixteen Australian university health sciences lecturers were the participants in the study, nine in the intervention group and seven in the control group. The teaching development carried out as part of the study had four inter-related objectives.

- a. To improve teaching skills and practices;
- b. To increase lecturers' focus on students and their learning;
- c. To improve student learning; and
- d. To assist lecturers to reach specific teaching goals.

In addition to the sound research design, the impacts of the programs were determined through the use of a number of indicators and measures incorporating quantitative data on teaching skills and qualitative data on teaching and learning quality. Data from the following were examined.

- A student evaluation of teaching questionnaire noted for its reliability and validity
- A teaching staff self evaluation questionnaire, the dimensions of which matched those of the student evaluation questionnaire
- A teaching orientation and focus questionnaire
- Student learning outcomes
- Entries in a journal kept by each teacher in the intervention group
- A treatment package effect measure.

In order to attempt to separate out the effects of the various components of the intervention that was undertaken in the study, a treatment package effect measure was employed at the end of the study. In essence, this questionnaire asked participants in both groups to rank the elements of the intervention and study of which they were aware. The questionnaire for each of the groups was slightly different as the intervention group had thirteen elements listed and the control group had seven listed. The idea of a treatment package effect is that the effect(s) of the various elements of the 'treatments' are determined. The broad outline of the original study design, including the indicators used, is presented in Table 1.

The original research was conducted in the Faculty of Health Sciences at La Trobe University in Victoria, Australia, between 2005 and 2006. La Trobe University was established in 1967 and is a large suburban and regional university geographically dispersed across seven campuses. Its faculties offer a wide range of professional and generalist courses, combining classic disciplines with professional and technical fields (Martens & Prosser, 1998). Participants were recruited in late 2004 and early 2005, and data were collected in 2005 and 2006.

The current study took place in the context of various national and institutional reforms and changes around teaching and learning in universities in Australia. While these are not detailed here, it is noted that, "Educational constructs, like those in other social sciences, are...complex, consisting of an array of contextual factors which can interact with each other and the variables under study" (Kember, 2003, p.94). As mentioned, the original study examined the impacts of a particular method of teaching improvement. Hence a control group was employed to take into account the impacts of the national and local reforms that were occurring at the time the study was conducted.

Table 1: Study design

Time	2005 (T1)	2005-2006	2006 (T2)
Group	Pre-intervention data examined	Individual Program Intervention?	Post-intervention data examined
Intervention Group (9 lecturers)	Quantitative student evaluation of teaching data Qualitative student evaluation of teaching data Self evaluation of teaching data Teaching foci questionnaire responses Student assessment results Treatment Package Effect responses	Yes	Quantitative student evaluation of teaching data Qualitative student evaluation of teaching data Self evaluation of teaching data Teaching foci questionnaire responses Student assessment results Treatment Package Effect responses Journal entries
Control Group (7 lecturers)	Quantitative student evaluation of teaching data Qualitative student evaluation of teaching data Self evaluation of teaching data Teaching foci questionnaire responses Student assessment results Treatment Package Effect responses	No	Quantitative student evaluation of teaching data Qualitative student evaluation of teaching data Self evaluation of teaching data Teaching foci questionnaire responses Student assessment results Treatment Package Effect responses

Challenges inherent in the study

There were at least two broad challenges inherent in conducting the original investigation. One comes from the general lack of evidence-based investigations of the outcomes of academic development (Kane, Sandretto & Heath, 2002). While there is some research in this area, the body of literature is small, and therefore, there is currently little to go on in terms of guiding development programs with evidence of what ‘works’.

Despite the increasing recognition of the value of teaching knowledge and skills in higher education systems worldwide, most university teachers do not have teaching qualifications or a high level of formal teaching development. In recent times, Australian universities have increasingly been involved in the development and provision of graduate certificates in Higher Education/Tertiary Teaching and the like and the Australian federal government appears keen to encourage activity in this area. But these programs and interventions operate in a climate of genuine resource and time constraints, as well as attracting some skepticism among university staff about the value of intensive teacher training. In one sense, this skepticism may be well placed. The questions of whether or not various teacher development interventions actually work and, if so, in what ways such interventions influence skills, practices, and foci, and/or ultimately lead to improved learning, remain largely unanswered in higher education.

In relation to group programs designed to improve teaching, Richard Johnstone, Executive Director of the national Australian Learning and Teaching Council, has noted that in relation to group programs such as graduate certificates,

[t]here are open questions about the extent to which these formal programs can be demonstrated to have material effect. The way you would do that would be to identify links between cohorts of people who've done these programs and the results in terms of evaluations of their own teaching.... Individual universities have done small studies but studies have not been done on a systematic basis in Australia (Devlin, 2006, p. 9).

In relation to individual programs designed to improve teaching, the body of research around the use of intervention with individuals is small, "... in some cases only peripherally related to the intervention and to date not at all assessing the effects of consultation on student learning outcomes" (Weimer & Lenze, 1997, p. 221). The original study sought to contribute to the research in this area.

A second challenge inherent in the investigation under discussion in this paper comes from the difficulties inherent in 'measuring' teaching effectiveness. A number of questions are central. How should 'effective teaching' be understood? What instruments or methods should be used to best determine whether teaching is effective, or has become more effective since an academic intervention took place? To what extent do reliability and validity of these measures matter?

The current paper grapples with some of the issues and questions inherent in these two challenges. In particular, it examines the attempts in the study to determine whether or not teaching had improved as a result of an intervention.

An understanding of 'effective teaching'

In order to determine what constituted *improved* teaching, *effective* teaching was defined in the original study. Any improvement or development in teaching was then clearly identified as movement toward this particular understanding of effective teaching. The scope of the current paper does not allow a full articulation of that definition. However, in

essence, for the purposes of this paper, effective teaching can be understood as teaching that

- incorporates appropriate teaching skills and practices
- is oriented to and focused on students and their learning
- facilitates appropriate student learning outcomes
- meets the particular requirements of the context in which it takes place.

Results of the original study

Overall, the study indicated that there is some merit in the particular approach to individual teaching development it employed. The study sought to determine the effects of an individual teaching development program on lecturers' teaching skills/practices; lecturers' focus on students/learning; student learning outcomes; and lecturers achieving specific teaching goals. Berk (2005) proposes that evidence should be collected from a variety of sources to both define teaching effectiveness and to determine whether or not such effectiveness is apparent. Berk argues that drawing on a number of sources of evidence allows the strengths of some to compensate for the weakness of others, thereby allowing a more robust decision about teaching effectiveness.

The results of the study under discussion show that the intervention employed was somewhat effective in improving specific aspects of the quality of participants' teaching, as measured by a number of indicators. While not every source of data about each of the four areas provided unequivocal evidence of the effectiveness of the intervention, there is some evidence of the effectiveness of the approach overall. As Kember (2003) puts it, "...the aim [was] ... of establishing a claim beyond reasonable doubt rather than absolute proof or causality" (p. 97). Kember (2003) further suggests that, "If a number of types of evaluation seeking data from multiple sources indicates that there was a measure of improvement in the targeted outcome, it seems reasonable to conclude that the innovation was effective" (p. 97). The targeted outcome in this case was an improvement in the teaching of the intervention group participants through the use of a particular individual teaching development program. It would seem reasonable, on the basis of the data collected in the study, to conclude that the individual teaching development program used was somewhat effective in improving teaching.

Statistical results

Statistical tests found no significant differences between the intervention and control groups in terms of pre- and post-intervention gains on the student evaluation of teaching instrument scales examined. A lack of statistically significant differences between the control and intervention groups is often taken to mean that the intervention has made no difference and in the case of the study under consideration, one straightforward interpretation would be that the intervention used did not make a significant difference to teaching effectiveness. However, the qualitative evidence, as well as the direction of some

of the other quantitative findings, would seem to challenge such a straightforward interpretation.

Further, there are at least three other possible explanations for the lack of statistical significance in the differences found in the study. One possible explanation is the small sample size. Given the very small samples of teachers in each of the intervention and control groups ($n = 9$, $n = 7$, respectively), the absence of statistically significant differences between the two groups is not unexpected. As part of their examination of methodological reasons for modest effects of feedback on teaching, L'Hommedieu, Menges and Brinko (1990) recommended that stratified random assignment and covariance analyses be used in conjunction with a sufficiently large number of teachers so that the initial equivalence of the groups is ensured.

The number of teachers in the study was not large enough to ensure such equivalence and it was evident from the demographic data collected that such equivalence may, indeed, have been absent. For example, on the whole, the control group consisted of less experienced teachers than the intervention group (an average of 11.4 years compared to the intervention group's 13.8 years of higher education teaching experience at pre-intervention). Further, on the whole, members of the control group were appointed at a lower level, with seven Lecturer (level B) appointments in the control group compared to the intervention group's four Lecturers and five Senior Lecturers (level C). Further, one member of the intervention group became an Associate Professor (level D) and another was an Acting Head of School during the study. These differences may have meant that control group participants as a whole were at an earlier stage of teaching development and more open to suggestion and change and/or that they had less responsibility and, therefore, more time to spend on teaching than the intervention group. Whatever the influences of the differences, the small sample sizes mean that individual and group differences could well have impacted on the results in ways that contributed to the absence of statistically significant differences between the groups.

In addition, the number of teachers in the present study was not large enough to absorb any perverse influence from the data related to one or more of the participants. It is likely that at least one control group participant had such an influence. This participant had a significant and very negative experience with her students the week before the students' evaluations of her teaching were collected in the pre-intervention stage. Specifically, in response to their continuous talking during lectures she stopped a lecture and pointed out how difficult it was for her to 'talk over' the students. She also pointed out that as an overseas academic she compared them to other students she had taught and that, in her view, they were 'letting the side down'. She further suggested to them that the continual talking would be foremost in her mind when answering the question, 'What are Australian students like?'

While a small number of students indicated that they were grateful for the lecturer dealing with the talking during class because it bothered and distracted them, many members of the class objected strongly to the lecturer's comments and manner in communicating those comments to the class. Objections were voiced to the lecturer directly at the time

she made them and then were evident in student comments on the student evaluation of teaching instrument, which was administered at the beginning of the next lecture. Those students who objected stated that they believed the lecturer was overgeneralising, could better handle the disruption and could be friendlier than she appeared to be. More than sixty comments in relation to characteristics of this lecturer's teaching that students perceived were important for her to improve in 2005 related to this event in the previous lecture and many revealed strong student objections to the lecturer's comments and perceived manner in communicating them.

With the absence of any such event in 2006, this lecturer had the largest gains in post-intervention student evaluation of teaching instrument scores of any participant in the study and the gains were, relatively, very large. Despite only being one person, the effects on the average scores of the very small control group from this single participant are likely to have affected the overall results of the study.

It is also possible that the circumstances of at least two of the intervention group may have had perverse influences in one way or another that could not have been absorbed because of the small sample size. For example, one participant was an Acting Head of School with significant responsibility outside of teaching. Having taken on considerable higher and extra duties, he deliberately chose only one goal area for improvement and was likely to have had less opportunity to integrate suggestions into his teaching practice during the period of the study. In fact, he made comments to the researcher toward the end of the study that reflected his concern that this may have been the case.

In the second example, the subject in which an intervention participant taught employed different modes of delivery at T1 and T2 – specifically, the subject was taught in block (intensive) format in 2005 and in evening format over an entire semester in 2006. This may have had a particular effect on the type of students who chose the subject, on the participant's approach to teaching and/or on her students' experience of learning in the subject that may well have limited any actual or perceived teaching improvements. Again, in such a small sample, this one participant's circumstances could have had an influence on the results.

A second possible reason for the absence of a statistically significant difference between the intervention and control groups on the student evaluation of teaching instrument scale scores is that the *John Henry* effect may have been evident. The *John Henry* effect is about compensatory rivalry by the control group. Commenting on their own study of the effectiveness of feedback and consultation with pre- and post-test measures, Marsh and Roche (1993) conclude that, "...we suspect that the act of volunteering to participate in the program, completing self-evaluation instruments, ...and trying to obtain positive SETs...may have led to improved teaching effectiveness of control teachers that reduced the size of experimental/control comparisons" (p.248). (SETs are student evaluations of teaching.

Given that one of the reasons for volunteering to participate in the study was that some of the control group teachers wanted to document their teaching effectiveness for promotion

purposes, it is possible that their desire to improve their teaching, coupled with the reflective self-evaluation experiences and specific, detailed student feedback data may have led to teaching improvement among control group teachers that would have reduced the difference between the control and intervention groups.

In the study under consideration, as implied by their volunteering for the study, continuing their participation for a year after being placed in the control group, and completing the self evaluations and other questionnaires at T1 and T2, members of the control group were keen to improve their teaching. However, they were left 'on their own' and may, therefore, have been particularly determined to improve their teaching. On the other hand, the evidence from the low number of requests from control group participants for materials available to all participants from a solution bank seems to suggest that although it may have had some effect, the *John Henry* effect was not prevalent in the study.

A third possible reason for the absence of statistically significant differences between the intervention and the control groups is the presence of a ceiling effect. McKeachie *et al.*, (1980) found that the effects of an individual program were most helpful to teachers with the lowest initial SET ratings. Overall in the study under consideration, the initial quality of teaching was high in both groups. This left less room to move, or more specifically, less room to improve teaching. Within the small window open in terms of room for improvement in the study, a statistical difference between the two groups would be very difficult to obtain. Further, given the high pre-intervention ratings by students on the student evaluation of teaching instrument, it is possible, as Piccinin (1999) suggests, that any improvement in teaching is not as readily perceived by students as it might have been if the starting point had been one where teaching improvement was clearly necessary. Where there was a large shift from pre- to post-intervention student evaluation of teaching instrument scale scores for one control group participant, as mentioned above, this gain came from a relatively low starting point.

It is likely that these three factors – the small sample size, the possible *John Henry* effect and the presence of a ceiling effect – may have, individually or in combination, contributed to the absence of statistically significant differences in the results of the study.

Research design

A number of aspects of the research design of the study warrant specific comment. They are discussed here.

The effects of voluntary participation

The study used voluntary participants. The major drawback to voluntary participation is that it limits the generalisability of findings (Gay & Airasian, 1992). However, there are also some benefits from voluntary participation that are particularly relevant here. Participants in the study agreed to have their teaching examined closely, to receive feedback and, if placed in the intervention group, to work collaboratively with the

researcher to improve their teaching. Brinko's (1993) review of the literature on the practice of giving feedback to improve teaching concluded that in a dozen studies in the areas of education, psychology and organisational behaviour, volunteers were found to be the most receptive to feedback.

In the study under discussion, the intervention group participants eagerly took on board suggestions that they pay attention to how the teaching was affecting their students and their students' learning. It may be that because they were voluntary participants, there was less likelihood of such reluctance.

Marsh and Dunkin (1992) refer to a sampling issue that impacts on generalisability and is inherent in using voluntary participants in a study such as the one being considered. Specifically, they point out that teaching staff who volunteer for a research project may be more highly motivated to improve their teaching than staff who 'naturally' seek out individual consultation at a university academic/teaching development centre. And these voluntary staff are also likely to be more motivated than those staff who are referred or compelled to attend such consultations because of poor teaching performance.

In the study under discussion, voluntary participation meant that it could be assumed that the participants had some level of curiosity about and a commitment to improving their teaching. This is likely to have created a platform for change that may have been a necessary component of the success of the intervention. A dependence on cooperation and/or a willingness to embrace change might limit the applicability and success of the approach with lecturers who may not have a choice in whether or not to take particular steps to improving their teaching (Devlin, 2003).

Piccinin, Christi and McCoy (1999) point out that it is not known whether the samples used in many of the outcome studies undertaken in the area of individual teaching consultation are representative of the population of staff who would typically use consultation services. Further, after spending some time examining the Australian and New Zealand higher education staff development units practices in the 1970s, Goldschmid (1978) noted that "...the observation is often made ...that many of those who seek advice ... are among the best and most concerned teachers and possibly need help less than the others" (p.234). However, it is possible that the 'typical' group who use individual consultation services in a university may be quite heterogeneous and that the results of the study are only applicable to those with a genuine interest in teaching who volunteer for individual consultation.

In any case, it is difficult to imagine an effective 'non-voluntary' individual teaching development program. While some universities make the use of teaching development services compulsory for teachers who are performing poorly, benefiting from an individual teaching development program can only really occur with an individual lecturer's voluntary cooperation in the improvement process.

External validity

External validity refers to the degree of generalisability beyond the sample involved in the investigation. Questions arise as to how well the intervention used in the study might be transferred to other academic development contexts outside those of the study, and how well it might be employed by other academic developers besides the researcher who undertook the original study.

Given the nature of the context-specific research in this investigation, in order for others to decide the extent to which these findings might relate to their own university setting, a number of sources of information are likely to be helpful. More specifically, the provision of detail about the intervention implemented, and rich descriptions and discussion of the application of the approach used, are most likely to be helpful in making decisions about the likelihood of successful transferability of the approach from the contexts described in the present study to other contexts.

It is possible that a number of uncontrollable external variables may have confounded the results of the study. However, these were managed by conducting the study in a single university and faculty where disciplinary differences were fewer than in a cross-institutional context; the inclusion of a control group; conducting the research over time to allow the novelty of participation to wear off and decrease the likelihood of pre-test sensitisation; and taking specific measures to ensure the treatment integrity of the study.

However, there was one other aspect of the research design that warrants further exploration – the control group treatment.

Control group treatment

The study under consideration used a research design that incorporated some ‘treatment’ for the control group. Specifically, the control group received student evaluation of teaching feedback between T1 and T2 and undertook a detailed self evaluation of their teaching. Both this feedback and the self evaluation process are likely to have led to control group participants scrutinising their teaching more closely than would have been the case in a no-treatment control group. While this design was deliberately used to isolate, as far as possible, the effects of the particular development program utilised, the feedback and self evaluation formed an intervention of sorts for members of the control group. As mentioned earlier, control group participants were also allowed access to a ‘solution bank’, which was full of tips about how to improve teaching. Access to such repositories is commonly available to those interested in improving their teaching. However, for the two of the seven control group participants who chose to access and use the solution bank, there were likely to be further intervention effects. Future research into the effects of teaching development should ideally include a number of control groups with varying levels of intervention in order to more specifically identify the components that have the most impact on teaching improvement.

Future research

In addition to considering differential control group treatments, future research might also benefit from employing a longitudinal approach.

One area highlighted by the study that would provide possible future research paths would be the introduction of a longer delay in the research design. As Gibbs and Coffey (2004) note, there can be "...a delay before changes in teachers'...teaching can significantly affect their students'... [learning]" (p.97). Piccinin, Christi and McCoy (1999) found a delayed effect in teaching improvement as measured by student ratings of teaching. They concluded that such a finding "...clearly points to the importance of tracking changes in teaching performance over time..." (p. 84).

Initiatives aimed at improving teaching can be deemed to have 'worked' when they improve student learning outcomes. It is essential that information that shows that the learning experiences of students taught by academic staff are improving be gathered. This is easier said than done but this should not dissuade efforts to begin what is likely to be a lengthy process. Ramsden (2003) estimates that changes to teaching can take between five to ten years to provide evidence of improved student learning experiences. The study under consideration sought to gather such evidence after just one year.

Guskey (1986) posed a model in which changes in classroom practice precede changes in student learning outcomes and the evidence of the latter change brings about changes in teaching beliefs and attitudes. It may be that there was insufficient time within the scope of the study for the evidence of student learning outcomes to affect teaching focus, one of the indicators of teaching effectiveness and improvement used in the study under consideration, and with greater time, the gains in this area could be further increased.

A longitudinal research design that provided the opportunity for changes to be made to teaching and to filter through to student learning would also be helpful in determining whether the broad positive changes evident in the results of the study are maintained, and possibly increased, over time.

Methodological rigour

There is little doubt that higher education research, like many areas of research, would generally benefit from increased rigour overall. Oakley (2003) suggests changes to the ways in which educational research is conducted and reported are necessary so that "...much more of it can enter the field of usable, robust evidence" (p. 1). While unable to address the myriad of challenges and issues related to conducting educational research, the study discussed in this paper represented an attempt to adopt as much rigour as was possible within the scope of the investigation. The use of a pretest-posttest control group intervention design with random allocation as well as the use of a psychometrically sound student evaluation of teaching instrument went some way to addressing some of the research-related issues that are well known and have been raised by Oakley (2003) and

others. More specifically, the present investigation utilised a sound method and an appropriate study type for the research questions it sought to answer. Future research should further build on these strengths and attempt to continuously improve the quality of the "...weight of evidence..." of its contribution to understanding in the field (Oakley, 2003, p.4).

As mentioned above, one way in which the rigour of the present study was ensured was through the use of a psychometrically sound student evaluation of teaching questionnaire. The student evaluation of teaching instrument used recognises the multidimensionality of teaching and has been developed through an extensive process involving the generation of an item pool from a literature review, forms in usage, interviews with university teaching staff and students, an examination of open-ended comments from students, ratings of the importance of the items in the pool, staff judgements on the items and the use of psychometric properties (Marsh & Dunkin, 1993; Marsh, 1994).

The use of such an instrument is somewhat rare in Australia – as Devlin (2004) notes, in relation to the typical process of development of student evaluation of teaching questionnaires.

It is not uncommon for a number of staff in a university to contribute suggested items or questions to a bank, from which some or all may be drawn to make up an instrument. The items or questions may be related to the teacher, the subject/course, the environment, facilities, resources, the provision of ICTs and any other factors in any combination. Often, the measure of an element of the student's experience is from a single item or question, rather than a scale containing a number of items or questions. Items, questions and whole instruments are rarely piloted and normative data almost never compiled (p.136).

Further, the student evaluation of teaching instruments that result from such processes are often unidimensional in terms of measuring teaching effectiveness. Yet instruments that recognise the multidimensionality of teaching are crucial. This is because, as Marsh and Roche (1993) note, "... teachers vary in their effectiveness in different SET areas as well as in their perceptions of the relative importance of the different areas, and that feedback specific to particular SET dimensions is more useful than feedback on overall or total ratings or feedback provided by SET instruments that do not embody this multidimensional perspective" (p.249).

Numerous comments from participants in the study under consideration in their journal and on the treatment package effect questionnaire confirm the usefulness of multidimensional feedback in terms of providing specific, focused information about particular and specific aspects of teaching. The SET instrument used was also reliable and valid.

Conclusion

Using one recent study of the effectiveness of a particular approach to assisting individual university lecturers to improve their teaching (Devlin, 2007), this paper has examined a

number of issues related to measuring teaching effectiveness. These include the challenges inherent in studies of this nature; the need for careful analysis of statistical results; consideration of aspects of the research design that might impinge on results, such as the use of voluntary participants, external validity and the control group treatment; and methodological rigour. The paper has also made several suggestions for future research of this kind. It is hoped that this paper may be of interest to others conducting educational research on the impact and/or effectiveness of interventions designed to improve teaching and learning.

References

- Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17(1), 48-62.
- Brinko, K. T. (1993). The practice of giving feedback to improve teaching. *Journal of Higher Education*, 64(5), 574-593.
- Devlin, M. (2003). A solution-focused model for improving individual university teaching. *International Journal for Academic Development*, 8(1-2), 77-89.
- Devlin, M. (2004). Communicating outcomes of students' evaluations of teaching and learning: One-size-fits-all? In C.S. Nair (Ed.), *Refereed Proceedings of the 2004 Evaluation Forum: Communicating Evaluation Outcomes: Issues and Approaches*. Monash University, Melbourne, 24-25 November, 2004, pp. 132-140.
- Devlin, M. (2006). Teaching the teacher. *Campus Review*, 16(6), 8-9.
- Devlin, M. (2007). *An examination of a solution-focused approach to university teaching development*. Unpublished PhD Thesis, Centre for the Study of Higher Education, The University of Melbourne, Australia.
- Gay, L. R. & Airasian, P. (1992). *Educational research (7th edition)*. Upper Saddle River: Merrill Prentice Hall.
- Gibbs, G. & Coffey, M. (2004). The impact of training university teachers on their teaching skills, their approach to teaching and the approach to learning of their students. *Active Learning in Higher Education*, 5(1), 87-100.
- Goldschmid, M. L. (1978). The evaluation and improvement of teaching in higher education. *Higher Education*, 7, 221-245.
- Guskey, T. R. (1986). Staff development and the process of teacher change. *Educational Researcher*, 15(5), 5-12.
- Kane, R., Sandretto, S., and Heath, C. (2002). Telling half the story: A critical review of research on the teaching beliefs and practices of university academics. *Review of Educational Research*, 72(2), 177-228.
- Kember, D. (2003). To control or not to control: the question of whether experimental designs are appropriate for evaluating teaching innovations in higher education. *Assessment and Evaluation in Higher Education*, 28(1), 89-101.
- L'Hommedieu, R., Menges, R. J. & Brinko, K. T. (1990). Methodological explanations for the modest effects of feedback. *Journal of Educational Psychology*, 82, 232-241.
- Marsh, H. W. (1994). *Students' Evaluation of Educational Quality (SEEQ): A Multidimensional Rating Instrument of Students' Perceptions of Teaching Effectiveness*. Self Research Centre: University of Western Sydney.

- Marsh, H. W. & Dunkin, M. J. (1992). Students' evaluations of university teaching: A multidimensional perspective. In J. C. Smart (Ed.), *Higher education: Vol 8. Handbook on theory and research* (pp. 143-234). New York: Agathon.
- Marsh, H. W. & Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American Educational Research Journal*, 30(1), 217-251.
- Martens, E. & Prosser, M. (1998). What constitutes high quality learning and how to assure it. *Quality Assurance in Education*, 6(1), 28-36.
- McKeachie, W. J., Lin, Y.-G., Daugherty, M., Moffett, M. M., Neigler, C., Nork, J., et al. (1980). Using student ratings and consultation to improve instruction. *British Journal of Educational Psychology*, 50, 168-174.
- Oakley, A. (2003). The 'new' technology of systematic research synthesis: challenges for social science. Paper presented at the Education, New Technologies, Local and Global Challenges: Learning in new environments conference, Madison, Wisconsin, U.S.A.
- Piccinin, S. (1999). How individual consultation affects teaching. In C. Knapper & S. Piccinin (Eds.), *Using consultants to improve teaching, New Directions for Teaching and Learning* (Vol. 79, pp. 71-83). San Francisco: Jossey-Bass.
- Piccinin, S., Christi, C. & McCoy, M. (1999). The impact of individual consultation on student ratings of teaching. *The International Journal for Academic Development*, 4(2), 75-88.
- Ramsden, P. (2003). Chapter 1: Introduction. In *Learning to teach in higher education (2nd edition)* (pp. 3-13). London: RoutledgeFalmer.
- Weimer, M. & Lenze, L. F. (1997). Instructional interventions: A review of the literature on efforts to improve instruction. In R. P. Perry & J. C. Smart (Eds.), *Effective Teaching in Higher Education: Research and Practice* (pp. 205-240). New York: Agathon Press.

Professor Marcia Devlin is Chair of Higher Education Research at Deakin University, Victoria, Australia. Her research involves theoretical and practical investigations into contemporary higher education issues, policies, practices and trends as well as university teaching and learning.
Email: marcia.devlin@deakin.edu.au