

Assessing the Impact of Teaching

John M. Malouff, Nicola S. Schutte, and Sally E. Rooke

University of New England, Australia

© 2008

Address correspondence regarding this article to John M. Malouff, School of Psychology, University of New England, Armidale, NSW, Australia. Email: jmalouff@une.edu.au.

Abstract

University teaching can have a positive impact, a negative impact, or no impact. This article describes indicators of the impact of university teaching at the unit level.

Teaching-impact indicators can be organized by the main beneficiary of the teaching: students; others, such as employers and clients, who interact with the students; the university; the department; and the instructors. Some teaching-impact indicators, such as student satisfaction with teaching, are commonly assessed. Other teaching-impact indicators, such as the percentage of students who complete a unit, could be assessed with routinely collected data. There are also teaching-impact indicators that require specific types of data, such as the results of knowledge tests of equivalent difficulty before and at the end of a course. Other types of impact data involve the effects of a course on individuals with whom the students interact in some way as part of a learning experience. Potential uses of teaching-impact data include evaluating instructors and improving teaching at the level of the individual instructor and across educational organizations and departments.

Assessing the Impact of Teaching

Traditionally, the assessment of teaching has focused on the quality of the teaching, including such elements as teaching style and methods. Over the past few decades there has been increasing assessment of the impact of government-funded activities such as mental health treatment and academic research. This development sets the stage for efforts to assess the *impact of teaching*. Theoretically, teaching can have a positive impact, a negative impact, or no impact. Information from other areas where impact is of concern provides context for examining assessment of teaching impact.

Information from the well-researched area of assessment of the impact of psychotherapy may provide insights into the possible trajectory of assessment of teaching impact. For the past 50 years or so university researchers have examined whether a treatment of some sort is statistically superior to no treatment or placebo treatment in producing mental health improvements. More recently, researchers have focused also on the *clinical impact* of treatment, sometimes called clinical significance (Jacobson & Truax, 1991) or social impact (Kazdin, 2002). Clinical impact can have many aspects such as clients moving out of the clinical range into the normal range of functioning or clients improving enough to return to work or school. For instance, Stulemeijer, de Jong, Fiselier, Hoogveld, and Bleijenberg (2005) showed that a cognitive-behavioral treatment led students with chronic fatigue syndrome to return to school at a higher rate than students on a treatment waiting list. The various meanings of clinical impact share the idea of a practically useful improvement. At present, psychotherapy research journals tend to require researchers

to report on evaluation of clinical impact (e.g., *Journal of Consulting and Clinical Psychology*, 2007).

Focusing on research impact provides further insights into the assessment of impact. The concept of research impact is presently in vogue at some universities (see e.g., Australian Department of Education, Science and Training, 2005; Coates, 2007) because governments provide universities with huge sums of money for research and want that research to produce important outcomes (Kostoff, 1994; Nicholls, 2007). Research impact can be measured in different ways, all sharing the concept of practical benefits, such as the use of findings to lead to improvement in government policies and practices or to improved medical treatments (Australian Department of Education, Science and Training, 2005). A fascinating example of research impact involves Barry Marshall, an Australian medical researcher who strove to show that a specific bacterium caused most duodenal and gastric ulcers. After years of effort, Marshall convinced the medical world that he was right and later received the Nobel Prize in Medicine or Physiology for his work (Nobelprize.org, 2005). His discovery led to new, more effective treatments of ulcers. Assessment of research impact may eventually guide funding of universities, departments, individual academics, and proposed projects (see e.g., Australian Department of Education, Science and Training, 2005). It could also play a role in university hiring and promotion.

As assessment of practical impact has become more common in psychotherapy research and begun to be explored in the evaluation of academic research, some educators have looked beyond the quality of teaching to the practical impact of teaching. For instance, there have been efforts to evaluate *teaching impact* at the university level in the context of student achievement of “graduate attributes”. These efforts have largely taken the form of surveying graduates and testing students

for specific important skills before they enter a university and after (Bath, Smith, Stein, & Swann, 2004). There have also been suggestions about evaluating academics partly through the financial impact of the units they teach (Malouff, 2007). Evaluation of teaching impact may be a wave of the future because governments that invest huge sums of money in education will one day demand specific information regarding the impact of that education on students.

Little has been written about the concept of teaching impact at the level of the outcomes of an individual instructor or unit. Because instructors and the units they teach are the building blocks of a university education, evaluation of the impact of individual instructors and their units is important. The most potent ultimate use for this sort of micro-assessment of impact would be improvement of teaching (and subsequent increases in teaching impact). Important questions include: (1) What are indicators of teaching impact at the level of an instructor or unit? (2) How does one choose among the indicators for specific purposes? This article provides preliminary answers to these questions.

Indicators of Teaching Impact

Teacher evaluation at all levels of education usually involves evaluation of teaching performance, either by other educators or students (Shinkfield & Stufflebeam, 1995). At the university level, assessment of teaching often starts and ends with student evaluations of instructor performance. There is evidence that this method of assessment is subject to a “halo effect,” which involves the student’s evaluation of one attribute of the unit or the instructor influencing his or her evaluations of other components (Darby, 2007). This halo effect complicates any attempt to identify positive and negative aspects of the unit. However, overall, student evaluations seem to have reasonably high reliability (consistency) and validity as

measures of student satisfaction (Greimel-Fuhrman, & Geyer, 2003; Marsh & Roche, 2000).

Other indicators of teaching impact can be identified. Impact indicators vary with regard to who is the main beneficiary of the impact. At least five potential beneficiaries exist: the student, others such as employers who interact with the student, the university, the instructor's department, and the instructor. We will discuss below indicators relevant to each of these types of beneficiaries.

Impact on the Student

Impact of teaching on an individual student is perhaps the most important type of impact as other impacts of teaching are likely to be mediated by impact on the student. University students generally want to learn, to be entertained, to be treated well, to obtain credentials, and to obtain employment (or a higher level or different kind of employment). Possible indicators of teaching impact on students include (a) increased knowledge or skills, (b) increased career preparedness, (c) increased motivation to learn more in general, (d) increased motivation to learn more about the content of the unit, (e) feeling satisfied with the educational process, and (f) feeling satisfied with the educational outcome.

Increases in knowledge, skill, career preparedness and specific types of learning motivation can be assessed by comparing pre and post instruction results. Pre-instruction assessment, which is currently done only rarely, is essential to evaluating increases. The best measures for increases in student knowledge involve carefully created and previously validated achievement tests, but these are rare. As a practical matter, individual instructors or departments may have to create the knowledge and skills tests – a substantial undertaking, because the measure must be either the same pre and post or shown to be of equivalent difficulty. On the positive

side, the effort would allow one to determine how much the students know at the start of a unit and how much knowledge they add during the unit – information most instructors never acquire.

The amount of knowledge or skill increase may be attributable to various factors other than the instructor or specific unit characteristics. For instance, student ability and effort can play important roles.

Student satisfaction can be assessed with instructor and unit evaluations completed by students and by observations of class behavior by other instructors visiting the classroom for the purpose of evaluating the unit or instructor (see e.g., Arends, 2006).

Impact on Others

Looking for indicators of teaching impact raises an important question: What are the purposes of educating students? One purpose is to benefit others with whom a student interacts, such as future clients, employers, employees, and family members. If, for instance, a university unit leads to improvement in the value of a student's work in a job, the employer benefits. For example, one of our students learned in a behavior modification unit how to use prompts and reinforcement to modify behavior. He then started a community practicum and used that behavior modification knowledge to motivate a client in a residential program for persistently mentally ill individuals to leave his room and become involved in therapeutic activities. Paid staff members asked the student how he accomplished that, and he explained how to use prompts and reinforcement. In that case the impact involved both the individual client and possibly the program, if the staff members applied what the student explained. Another of our students in a similar unit learned about the "caring days" marital therapy technique, which involves the two individuals doing three things each day

that the other person has identified as showing caring toward him or her. After the unit ended, this student used the method to improve her stale relationship with her partner.

The benefits to others can occur during the unit, e.g., as the result of an assigned project (Malouff, 2004) or work-placement, or after the unit ends. Naturally, the more individuals who benefit and the greater the benefit for each, the greater we would judge the “impact.”

Tracing a benefit back to a specific unit can be difficult, however, because of overlapping content and skill development among units. An organization likely to benefit from the acquired knowledge of a student might be an organization providing a work placement or first post-graduate employment. Obtaining information about the contributions of a single unit might best be accomplished through surveys of these organizations, with questions asking about specific students and specific knowledge or skills targeted in the unit of interest. It would not generally be cost-effective to conduct such surveys regarding a single unit, but it could be possible in one survey to assess various units and entire programs. Other possibilities for evaluating impact might involve collecting data systematically from students about beneficial applications of their knowledge, as when students apply learned principles to help others and carefully evaluate the effects of the applications. We have used this method in Behavior Modification assignments (Malouff, 2004), always informing students that their grades depended on the quality of their applications, not at all on the outcomes. We gave this assurance to reduce the chances that students would inflate their reported level of success.

Impact on the University

Universities generally value student progress and graduation, as well as the money that comes with student enrolment in units. Possible indicators of teaching impact on the university include (a) a large number of enrolled (paying or funded) students in a unit, (b) a large number of students who pass a unit, (c) a high percentage of enrolled students who complete the unit, (d) a large amount of money university received for students in the unit (the amount may vary with the nature of the educational program containing the unit), (e) an enhancement of the university's reputation through offering the unit or increased student satisfaction due to offering the unit, and (f) a high percentage of students in the unit who graduate within an expected period of time. The first four indicators involve data routinely collected by universities. University reputation and satisfaction effects, which would be rarely collected for a single unit, could be assessed through surveys, for example, of potential new students or current students. The graduation percentage indicator, which might be most relevant for basic or introductory units or study preparation units, would require special student-tracking data collection for assessment of impact.

Impact on the Department

Departments generally are interested in having a substantial number of students in their units. Possible indicators of teaching impact on departments include (a) the number (or percentage) of students enrolling in related units after completion of the unit to be evaluated, (b) the number enrolling in the instructor's higher level units, and (c) the number changing their course of study to the course of study incorporating the unit. Departments can obtain this objective information from university databases, but it would involve tracking individual students who completed the unit of interest.

Impact on the Instructor

Instructors generally want to enjoy teaching, improve their teaching skills, and learn more about the topics they teach. Possible indicators of teaching impact on the instructor include (a) learning more about the content of the unit, (b) improving his or her teaching skills and practices (Cross, 1990), and (c) enjoying teaching the unit. Assessment methods could include self-reflection and self-evaluation (Squires, 2003; Withers, 1995).

Choosing Reliable, Valid Indicators for Specific Purposes

A primary purpose of evaluating teaching impact by individual instructors is to evaluate and ultimately to improve the performance of the instructor. For assessment by the university, objective measures (e.g., percentage of students completing a unit) or measures involving subjective ratings by students and peers would be better than self-ratings by instructors, which could be affected consciously or unconsciously by incentives for high impact. Although universities rarely consider teaching impact information beyond student evaluations of teaching style, the above taxonomy suggests other important indicators, including student increases in knowledge and skills, increases in work readiness, positive impact on others who interact with the student, increased motivation to learn more about related subjects, and high levels of student completion of the unit.

Evaluation can focus on the entire unit or on aspects of a unit. For example, students may evaluate the usefulness of aspects of a unit such as a case study assignment, problem-based learning or a particular text as well as the unit as an entity. Information regarding impact of aspects of a unit can help instructors make targeted changes for the future.

Universities and departments can evaluate indicators in an absolute sense (e.g., the mean student evaluation of the unit) or relative to similar units. Both types of

evaluation can provide valuable information about the performance of an instructor and a specific unit. With this sort of information for multiple instructors and units, universities can evaluate not just individual instructors but whole departments and programs.

Universities and departments can also set targets for specific indicators for specific instructors or across all instructors. Setting specific targets or goals to which individuals feel commitment tends to help improve performance (Latham, Soosan, & Locke, 1997; Locke & Latham, 2006).

Instructors can also use the indicators for self-assessment and self-improvement. Here again objective indicators have great value, but self-ratings can also be helpful. Using impact data to create a systematic improvement plan (e.g., Squires, 2003) could aid instructors who are doing well or not so well. To use teaching-impact information to improve teaching, academics will need to identify the teaching methods, teaching style elements, and types of unit structure and content that help produce the impact. This involves switching from a focus on impact to a focus on process.

Impact assessment by anyone (e.g., students, instructors, work placement supervisors) can be quantitative or qualitative. Quantitative assessment, such as students rating their satisfaction with aspects of a unit on a five point scale, has the advantage of relatively easy interpretability and allows comparison across units and ratings of aspects of a unit. Qualitative assessment, for example, the narrative reports of a colleague on an instructor's teaching strengths after having observed the instructor, has the possibility of providing new information and perspectives that might not be gathered through quantitative assessment. Often both quantitative and qualitative assessment can be used together to provide comprehensive information.

In assessing teaching impact, it will often be difficult to separate the teacher from the unit. This is true because the unit is at least partly an extension of the instructor. For many purposes, separating the two does not matter. For evaluating an individual instructor, comparisons of the impact of that instructor with the impact of other instructors teaching the same unit could be useful. For evaluating a unit, collapsing impact data across the instructors who teach the unit might be best.

Using Teaching-Impact Data

To illustrate possible uses of teaching-impact data, we will describe two different hypothetical assessment situations.

Situation 1

The head of a psychology department collects impact data as part of a program review and notices major contrasts between two non-required 300-level units, Personality and Interpersonal Communication. The Interpersonal Communication unit has significantly higher unit ratings from students, and the Interpersonal Communication instructor has slightly higher instructor ratings than the Personality instructor. Many student comments about Interpersonal Communication describe beneficial personal applications of the subject matter in interactions with children, spouses and co-workers. A few student comments about Personality indicate that it led students to understand themselves better. Pre and post knowledge tests show that students in both units increased substantially in knowledge about the topic during the unit.

University databases show that 131 students enrolled in the Interpersonal Communication unit. Only 97 enrolled in Personality. Of the students who enrolled in Interpersonal Communication, 80% completed the unit; 68% of the students who enrolled in Personality completed it.

Surveys of psychology students who graduated between 6 and 12 months ago show that 30% of them rate Interpersonal Communication as a unit that has been “very helpful” in their work. Only 4% so rate Personality. A survey of work supervisors of the students indicates that 60% rate the official description of Interpersonal Communication as likely to be “very helpful” for individuals doing the work of the student. Only 15% so rate Personality. Surveys of prospective psychology students show that the two units are equally appealing to them. The department head does not collect any evaluation data directly from the two instructors.

Integrating the most important findings, the department head might conclude that the Interpersonal Communication unit (a) is more popular with existing students, but not with potential students, (b) has higher student-retention and student satisfaction and seems to lead to greater benefits for others who interact with the student, and (c) is considered more valuable by work supervisors of psychology students. If other comparisons of units show a similar pattern favouring practically useful units, the department head might (a) try to add units that have practical value, (b) discuss with the Personality instructor whether the instructor can make the unit more applied or practically useful, (c) encourage other instructors to make their units as practically oriented as possible, (d) present some of the non-confidential Interpersonal Communication impact data to students so they can make more informed choices about what units to take, and (e) praise the Interpersonal Communication instructor for teaching a unit with such positive impact.

Situation 2

A university promotion board is considering whether to promote a law instructor. The instructor has provided impact data that shows that (a) her teaching, across units, receives high instructor and unit evaluations, (b) she recently received an

award from the student law society for outstanding teaching, (c) a peer evaluation of her unit describes outstanding rapport with students and excellent student participation in class sessions, (d) her non-required units have unusually high enrolment rates, (e) her units are the most mentioned law units in surveys of graduates as “very helpful” in both their practicum work and in their first year of legal practice, and (f) surveys show that more students switched their main interest to her specialty area, Environmental Law, between entry and graduation, than with regard to any of the other 10 specialty areas, except one. With regard to teaching, the board might conclude that the instructor is having excellent impact and consider that in reaching a promotion decision.

Conclusion

Assessing the teaching impact of individual university instructors and units could enhance the quality and outcomes of higher education in several ways. Perhaps most important, knowledge of the proximal and distal effects of teaching content and methods used in individual units could help instructors improve the units. Information regarding instructor and unit impact can be used for university decisions relating to instructor promotion or other types of incentive schemes. Information relating to which units have high impact or have specific aspects with high impact can be used in developing new units or revising existing units. Where the number of units offered by a university is restricted by factors such as government funding, units producing higher teaching impact could be given priority. Information regarding teaching impact may also be used in government mandated quality assurance reviews. Finally, attending to practical impacts of teaching could ultimately have the valuable effect of enhancing the prestige of university teaching, which tends to compete with

research activities.

References

- Arends, R. I. (2006). Summative performance assessments. In S. Castle and B. D. Shaklee (Eds.), *Assessing teacher performance: Performance-based assessment in teacher education* (pp. 93-123). Lanham, MD: Rowman & Littlefield.
- Australian Department of Education, Science and Training. *Research quality framework: Assessing the quality and impact of research in Australia*. Retrieved December 7, 2007, from http://www.dest.gov.au/NR/rdonlyres/E32ECC65-05C0-4041-A2B8-75ADEC69E159/4467/rqf_issuespaper.pdf
- Bath, D., Smith, C., Stein, S., & Swann, R. (2004). Beyond mapping and embedding graduate attributes: Bringing together quality assurance and action learning to create a validated and living curriculum. *Higher Education Research & Development, 23*, 313-328.
- Coates, H. (2007). Excellent measures precede measures of excellence. *Journal of Higher Education Policy and Management, 29*, 87-94.
- Cross, K. P. (1990). Making teaching more effective. *Journal of the Freshman Year Experience, 2*, 59-74.
- Darby, J. (2007). Are course evaluations subject to a halo effect? *Research in Education, 77*, 46-55.
- Greimel-Fuhrman, B., & Geyer, A. (2003). Students' evaluation of teachers and instructional quality – analysis of relevant factors based on empirical research. *Assessment & Evaluation in Higher Education, 28*, 229-238.

- Jacobson, N., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12-19.
- Journal of Consulting and Clinical Psychology (2007). *Instructions to authors*. Retrieved December 4, 2007, from <http://www.apa.org/journals/ccp/submission.html>.
- Kazdin, A. E. (2002). *Research design in clinical psychology* (4th ed.). Boston: Allyn & Bacon.
- Kostoff, R. N. (1994). Federal research impact assessment: State-of-the-art. *Journal of American Society of Information Science, 45*, 428-440.
- Latham, G. P., Soosan, D., & Locke, E. A. (1997). Implications of goal-setting theory for faculty motivation. In J. L. Bess (Ed.), *Teaching well and liking it: Motivating faculty to teach effectively* (pp. 125-142). Baltimore: Johns Hopkins University Press.
- Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current Directions in Psychological Science, 15*, 265-268.
- Malouff, J. M. (2004). Undergraduates can learn about behavior therapy by using it to help others. *Behavior Therapist, 27*, 71-71.
- Malouff, J. M. (2007). *The Academic Productivity Index*. Manuscript submitted for publication.
- Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology, 92*, 202-228.
- Nicholls, M. G. (2007). The development of a benchmarking methodology to assist in managing the enhancement of university research quality. *Higher Education*

Quarterly, 61, 539-562.

Nobleprize.org (2005). *The Nobel Prize in Physiology or Medicine 2005*. Retrieved

December 4, 2007, from

http://nobelprize.org/nobel_prizes/medicine/laureates/2005/press.html.

Shinkfield, A. J., & Stufflebeam, D. (1995). *Teacher evaluation: Guide to effective practice*. Boston: Kluwer.

Squires, G. (2003). *Trouble-shooting your teaching: A step-by-step guide to analysing and improving your practice*. London: Kogan Page.

Stulemeijer, M., de Jong, L. W. A. M., Fiselier, T. J. W., Hoogveld, S. W. B., & Bleijenberg, G. (2005). Cognitive behaviour therapy for adolescents with chronic fatigue syndrome: Randomised controlled trial. *British Medical Journal*, 330, 7481-7485.

Withers, G. (1995). Getting value from teacher self-evaluation. In A. J. Shinkfield & D. L. Stufflebeam (Eds.), *Teacher evaluation: Guide to effective practice* (pp. 261-271).