

Evaluating University Teaching and Learning

Patricia J. Rogers
Royal Melbourne Institute of Technology, Australia
Email: Patricia.Rogers@rmit.edu.my

ABSTRACT

This paper discusses ways to evaluate university teaching and learning in ways that improve the human capital of teachers and students. While evaluation of teaching and learning often focuses only on gathering and using student feedback with evaluation, there is of course a wide array of literally dozens of evaluation models, approaches and tools. These many alternatives offer the opportunity to match each particular situation with the most appropriate approach, providing there is some way of reviewing options and selecting the most relevant one. This paper sets out some of these different approaches to evaluation, and a guide to selecting which approach to evaluation should be used when, and how these might be useful not only to judge the quality of teaching and learning but to actually improve it.

Introduction

In many universities the evaluation of teaching and learning often means only one thing – feedback from students on an individual course or subject. In Malaysia, Australia and other countries, considerable effort has gone into developing systems to gather and report feedback and use it to make subsequent improvements. While student feedback is obviously an important source of information, by itself it is likely to miss important aspects of teaching and learning. There is more to evaluation than this.

Most university teachers also gather informal evidence of student reactions and draw conclusions about their level of interest and

understanding. Where possible they engage students in questions and answers that indicate how well the concepts have been understood. Student assessment, whether through essays, exams, tests, practical exercises portfolios or reflections, also provides information about how well students have learned and, by implication, how well we have taught.

However these processes by themselves are insufficient. Student feedback may be ill-informed when students have not had an opportunity to apply their learning to a professional task. Observations of student behaviour in class can over-estimate or under-estimate how much students are learning. Assessments are unfortunately not always an adequate test of student learning and often come too late for remedial action.

There is also more to evaluate than individual subjects and courses. It can be important to evaluate specific interventions, such as textbook, methods of delivery such as webinars, aspects of the university experience which affect teaching and learning, such as access to support facilities and broad strategies such as language strategies. In this paper, I use the word 'intervention' to be clear that the discussion is not only about evaluating subjects and courses.

There are many ways of going about evaluation more comprehensively and systematically. Some focus on particular research designs or data sources for collecting evidence of learning such as standardised tests or randomised control trials and others focus on particular processes for engaging teachers, students and other stakeholders in both generating and using evidence about learning. Some approaches focus on improving teaching and learning for specific current students while other are primarily focused on building knowledge to improve teaching and learning in the future.

The paper provides a way to select the appropriate approach to evaluation by focusing on how evaluation is expected to improve university teaching and learning. The paper begins by setting out some different types of evaluation in terms of their purposes and key questions. It then discusses how evaluation can be used not only to judge the quality of teaching and learning but to actually improve it. It asks "If evaluation is part of the solution, what is understood to be the problem?" and sets out six different types of problems that evaluation might help to solve. Finally, it describes some very different approaches to evaluation that show how a good understanding of the purpose of the evaluation, and the way it is intended to improve teaching and learning, can help develop an appropriate evaluation for the situation.

Formative and Summative Evaluation

It can be helpful to begin by distinguishing between different types of evaluation. A fundamental distinction is between ‘formative evaluation’, aimed at improvement, and ‘summative evaluation’, aimed at selection (Scriven, 1967, 1991). Formative evaluation is more usually done during implementation and used by implementers to make incremental changes to what they are doing. Summative evaluation is more usually done either at the beginning of a project, to inform policymakers or funders in choosing between alternatives, or at the end, to inform the decision whether to continue, extend or terminate an intervention.

“[Formative evaluation] is typically conducted during the development or improvement of a program or product (or person, and so on) and it is conducted, often more than once, for the in-house staff of the program with the intent to improve.”

(Scriven, 1991, 169; emphasis in original)

“Summative evaluation of a program (or other evaluand) is conducted after completion of the program ... and for the benefit of some external audience or decision maker. ... The decisions it services are most often decisions between these options: export (generalize), increase site support, continue site support, continue with conditions (probationary status), continue with modifications, discontinue.”

(Scriven, 1991, 340; emphasis in original)

Stake (cited in Scriven, 1991) has summed up this distinction with the following metaphor: “When the cook tastes the soup, that’s formative evaluation; when the guests taste it, that’s summative.” The cook, after evaluating the soup by tasting it, can make adjustments (for example, adding more vegetables, or heating it more), whereas the customer, after evaluating the soup by tasting it, only has the option of continuing to eat the soup or rejecting it.

The metaphor is worth considering further, however. The cook might be seeking to improve the soup (formative evaluation), but decide it is so bad it should be thrown away (turning it into a summative evaluation). If many customers reject the soup (based on their summative evaluation), the cook may decide to change the recipe (turning it into formative evaluation). The customers might be able to take some actions based on their formative evaluation (such as adding salt or chilli sauce if these are

provided on the table). This shows that the concepts of formative and summative evaluation are more about how the information is used than about simply who uses it or when it is done.

So is student feedback on university subjects and courses, usually collected at the end of semester, formative or summative? It is usually intended to be formative – that is, used to improve the subject or course in the future. However, this means that any improvements to respond to the feedback will come too late to help those students who provided it. In my teaching I always schedule some feedback from students mid-way through the course (whether this is a semester-long course or a two-day short course), so that where possible I can correct problems before the end of teaching.

Five Types of Evaluation

Beyond this simple two-part classification, it can be helpful to clearly distinguish five different types of evaluation, building on the types identified by Rossi, Freeman and Lipsey (2000) and Owen (1992, 2007) as shown in the following table:

Table 1: Types of Evaluation Related to Purpose

1. Needs analysis	– to identify the needs that should be met
2. Design evaluation	– to set out a way to meet these needs
3. Process evaluation	– to document and review how implementation is proceeding
4. Impact evaluation	– to assesses results and the contribution of the intervention to achieving these

Before describing these in more detail, it is important to be clear that the evaluation of an intervention might need to include all or most of these to some extent.

For example, to plan the formative and summative evaluation of a multimedia package that was developed within the university to teach research methods to nursing students (Guice et al., 1996), the first four of these categories were used. We began with needs analysis, then developed a plan of how the intervention could be implemented in a way that would meet those needs, reviewed implementation, making changes as we went, and finally assessed the effectiveness of the package in

meeting the needs we had identified – and also the other impacts it had had.

Needs Analysis

Focus and Timing

Needs analysis focuses on answering the questions:

- What needs should the intervention meet?
- Which of these needs are currently being met?
- What should be the intended outcomes of the intervention?

Needs analysis is usually undertaken before an intervention is implemented, but may be done as part of developing the criteria for an impact evaluation.

Methods

It can be helpful to consider the following different types of need (Bradshaw, 1972): felt, expressed, normative and comparative.

Felt need refers to needs as expressed by the individuals themselves, usually gathered through direct data collection such as interviews and questionnaires. Sometimes felt needs are specific about the type of additional service or assistance that is required (such as better access to computer laboratories) and sometimes they simply identify a gap (such as needing some assistance in understanding a particular part of the course). Needs assessment can be done of participants' felt need before they begin learning, or of participants' felt unmet need after they have completed their learning. For example, a needs assessment was undertaken using questionnaire survey of junior doctors' knowledge and beliefs concerning evidence-based medicine. Their high ratings of both the importance of evidence-based medicine and their need for more training in it will inform the revision of doctor training for the future (Hadley et al., 2007).

Expressed need refers to needs that are expressed through visible evidence of demand. For evaluating university teaching and learning, this could include usage rate of particular services, or application rates for assistance.

Normative need is based on an expert's opinion of what is needed or agreed best practice recommendations. When evaluating university

teaching and learning, this could include national curriculum guidelines about what should be learned and how it should be learned. It could include employer surveys providing feedback on the skills and knowledge of university graduates or students on fieldwork placements. For example, the redesign of the podiatry course at Curtin University in Australia, (Kippen, 1995) included commissioning an independent needs analysis, collecting the opinions of those working in the industry, to identify the competencies that would be required by a podiatrist in the future.

Comparative need is based on an analysis of service levels compared to relevant benchmarks or similar organisations. This could include the level of resourcing such as teacher-student ratios, or access to specialist facilities such as computer equipment.

Tips and Traps

Only considering one type of need, and using only one source of evidence, raises the risk of missing important needs.

All approaches to needs assessment risk focusing on unmet needs only, ignoring needs which are currently being met and which should continue to be met. At its worst this can lead to courses being changed to leave out important elements, which then become identified next time as unmet needs.

Design Evaluation

Focus and Timing

Design evaluation focuses on answering the questions:

- How should the intervention be organised?
- How should it work in theory?

Like needs assessment, design evaluation is most usefully done as part of planning an intervention, but can be done retrospectively on an existing intervention to identify variables to include in data collection and/or provide a conceptual framework for analysis and reporting.

Methods

In many cases, it is useful to use some form of program logic or program theory (Rogers, 2000). A program theory, or logic model, is an articulated

model of how a program or project is understood or intended to contribute to its intended or observed outcomes. Some versions of program theory set out a fairly simple logic model as shown in Figure 1.

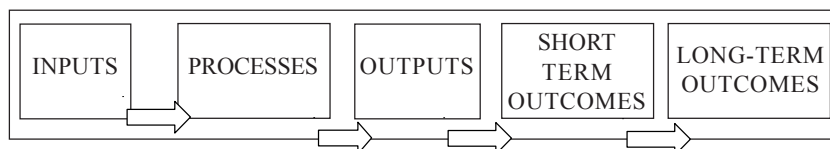


Figure 1: Simple Logic Model of Inputs and Outputs

This can be turned into a series of questions to guide the planning and evaluation of university teaching and learning, as shown in the following example drawn from the Center for Youth and Communities at the Heller School for Social Policy and Management, Brandeis Universities (NSLP, 2006).

[Insert Organization's Name Here]					
Template and Definitions for Logic Model					
For Whom	Assumptions (Theory of Change) Moving from...	Inputs	Strategies/Activities through...	Outputs	Outcomes through... Impact
Who are you ultimately doing this for? Who will benefit from your work and the changes you create? Who (described as specifically as possible) should this be in the first year? In two years? Three years? Four years?	Why are you doing this? Why does the program/organization exist in this local area? Why does change need to happen? What are the social, political and economic conditions or environmental factors that make you want to do it? What identified needs or assets led you to address the issue? What assumptions guide your work? What do you know, think, and believe about why you do what you do?	What resources will your program need to implement its activities?	What supports and opportunities are needed to provide a positive developmental setting for youth? What mix of programs, services, and activities will be needed to achieve the desired outcomes? What systems need to be in place?	What products or services were created and/or delivered by your program? How many?	What outcomes do you want to achieve for youth? What changes for youth do you want to see over the next four years and beyond? How would your community and institutions like to operate differently? Changes in the lives of the beneficiaries and/or participants that constitute significant long-term benefits to them.

Figure 2: An Example of a Template for a Logic Model

Others describe the program theory in terms of a chain of results, or an ‘outcomes hierarchy’ (Funnell, 2000). For example, Figure 3 shows an outcome hierarchy for a course in evaluation.

In either case, a key feature of program theory is that it articulates intermediate outcomes not a flowchart of tasks. While program theory has become increasingly popular in recent years, its use can be traced

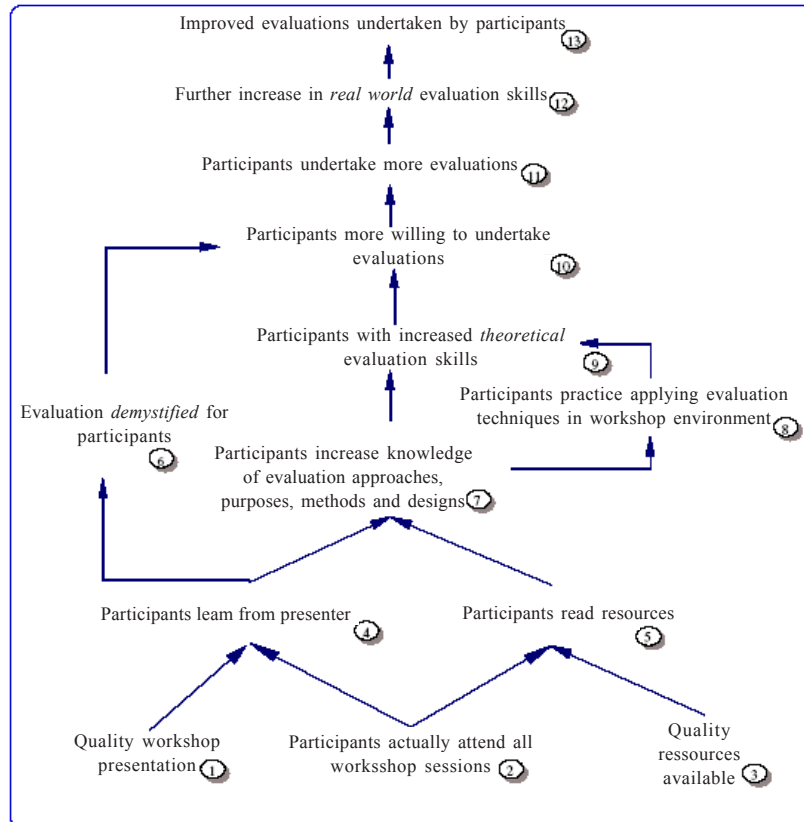


Figure 3: Example of An Outcomes Hierarchy (from Duignan, 2000)

back at least to the 1960s when Suchman (1967) suggested that evaluation might address the achievement of a ‘chain of objectives’.

It can be useful for planning purposes to work backwards from the intended long-term outcomes (particular learning outcomes) to identify what short-term outcomes are needed in order to achieve these. Information for developing a program theory for a university intervention can come from descriptions of similar programs, from research into the causal pathways by which students learn, or from the ‘practice wisdom’ of university teachers and others (such as employers).

Developing a program theory can be a very useful part of process and impact evaluation, as it can then form the basis of planning for data collection, analysis and reporting.

Tips and Traps

Without care there is considerable risk of developing a poorly conceived program theory that is not credible, and for this then to focus data gathering for the evaluation in a way that misses important evidence about what is happening and how results are being achieved (or not achieved). It is important therefore to open the program theory up to scrutiny by experts and to check it against what is known from research.

Another challenge in developing program theory is to create a diagram which is simple enough to be useful but detailed enough to adequately represent all the important elements in the intervention.

Process Evaluation

Focus and Timing

Process evaluation focuses on answering the questions:

- How is it going?
- How does it work in practice?

Process evaluation is ideally planned before implementation and carried out during implementation. It is sometimes possible to retrospectively conduct a process evaluation using interviews about people's experiences and documentation of implementation.

Methods

Process evaluation might focus on checking the quality of implementation, particularly if standards for activities have been specified. This can be used to improve implementation, where possible, and to interpret evidence from an impact evaluation. If an intervention is found to be ineffective, then it is critically important to know if it was properly implemented or not. Evidence for these types of process evaluations can come from observations (using standardised checklists and assessing the intervention for compliance).

Process evaluation can also be intended to carefully document an innovation so that, if it is found to be effective, it can be replicated or continued.

Process evaluation can also involve ongoing and regular review of progress and incremental changes to improve implementation. It is

important to not only look for matters that need to be changed but also to identify matters that need to be maintained.

In all types of process evaluation, evidence might include interviews with staff and students, program records (including performance indicators) and observation of implementation. If the evaluation is focusing on checking program quality, observation is likely to use a standardised checklist and judgements about whether it is complying with requirements or meeting benchmarks). If the evaluation is intended to document an innovation, observation is more likely to involve detailed descriptions.

Students may well be engaged in process evaluations not only as sources of information but as users of the information. Regular feedback on their learning can assist them to modify or continue their learning strategies, as much as it assists teachers to modify their teaching strategies. For example, Lee and Yeap (2005) described how web-based resources for university engineering education provided both students and lecturers with feedback on their progress, which helped them to improve their performance:

“Students can attempt the online Quiz section for a specific lecture topic. At the end of each quiz, the score was displayed along with suitable and immediate feedback that explained why the selected answer was incorrect. This ongoing feedback information is stored in the computer database and can be statistically analyzed by the lecturer to study the student preferred learning pattern of the course content and to determine the effectiveness of the questions from semester to semester. It also appeared that student achieved better marks in their overall performance when they were given with continuous improved e-notes and WebCT access.”

Tips and Traps

An area of concern in process evaluations is the appropriateness of criteria for assessing the quality of teaching. In comments directed to school education but also relevant for university education, Scriven (1990) has warned of the lack of validity of using variables that are correlated with student learning as proxy measures for evaluating the quality of teaching and as predictors of the outcome of student learning. Whether these variables are characteristics of teachers (which might erroneously be used to select teachers) or styles of teaching (for example, maintaining

eye contact with students), they are never perfect predictors of learning (in fact often not even highly correlated) and use of them would focus attention on complying with the teacher behaviour rather than on achieving student learning, and would penalise effective teachers who use other methods.

Dunkin (1997) discussed the importance of distinguishing between teacher effectiveness, teacher competence and teacher performance:

“Teacher effectiveness is a matter of the degree to which a teacher achieves desired effects upon students. Teacher performance is the way in which a teacher behaves in the process of teaching, while teacher competence is the extent to which the teacher possesses the knowledge and skills (competencies) defined as necessary or desirable qualifications to teach.”

Impact Evaluation

Focus and Timing

Impact evaluation focuses on answering the questions:

- What have been the outcomes?
- Did it work? For whom and in what ways did it work?

Methods

Impact evaluation requires both evidence of outcomes (student learning and other outcomes) and an analysis of the contribution of the intervention (a particular course or a teaching approach) to achieving these.

Evidence of outcomes might come from student assessment, employer surveys, exit interviews, feedback from students, and information about destinations of students (for example, acceptance into post-graduate courses).

Analysis of causal contribution can be done through using particular research designs such as experimental designs (randomised control trials) or quasi-experimental designs (such as use of a comparison group through matching or statistical construction of a comparison group through propensity scores or co-variance analysis).

Alternatively causal analysis can be undertaken by developing a series of testable hypotheses and testing these empirically by searching

for and explaining exceptions (participants who did not achieve the outcomes and non-participants who did), and by identifying and ruling out alternative explanations for the outcomes.

Tips and Traps

Some of the particular challenges in evaluating university teaching and learning relate to the time lag before long-term learning is evident. The real test of student learning will come when they try to use their knowledge in their profession after they complete their course. Not only is it more difficult to collect data after they have left, and to collect data about actual behaviour not just reported attitudes, it obviously comes too late to make timely changes to courses. One of the ways to address this is to use program theory to identify intermediate outcomes that need to be achieved. It may be possible to gather evidence of these more readily and to identify and address any problems.

The increasing international focus on the use of randomised control trials and quasi-experimental designs for impact evaluation increases the risk that alternative approaches will not be implemented rigorously when they are used, and not used when they are more appropriate.

How Evaluating University Teaching and Learning Can Improve It

Amongst these many choices about how to do evaluation, the most important principle is to make the selection of type and method based on an analysis of what is likely to actually make a difference in terms of improving teaching and learning. Table 2 below shows six different ways in which evaluation might be expected to contribute to improvements in teaching and learning. These are discussed in more detail below.

The ‘dashboard’ refers to the use of timely, credible data to make incremental adjustments to what we are doing. It is based on the metaphor of the car dashboard which provides the driver with real-time data about speed and temperature which can guide behaviour (sticking to the speed limits) or provide early warning of a problem (such as an engine overheating). Providing students with data about their learning, teachers with timely, relevant data about their students, or university administrators with similar data about courses, can be used to support incremental improvement and problem-solving.

Table 2: Different Ways of Improving Teaching and Learning Through Evaluation

Label	Condition	Pathways
1. Dashboard	Do not know if performance is good or bad	Get frequent, timely information on results, and use trial and error to improve
2. Advice	Know it's bad but do not know how to improve poor performance	Get knowledge about better ways to do it (through general advice or specific advice) and apply this
3. Copying		Learn from exemplars that can be copied
4. Support	Know it's bad and how to fix it but lack capacity to improve (resources or authority)	Generate support for adequate resourcing through advocacy
5. Carrots and sticks	Know how to fix it and can but lack incentives to make improvements	Increase the incentives to improve performance (either through the fear of sanctions or the promise of rewards)
6. Learning capacity	Cannot maintain or replicate good performance	Feedback on the outcomes of good performance, learn from success and how it can be continued or copied

A different strategy is needed when it is known there are problems but there is a lack of knowledge about how to improve performance. Advice from mentors or peers may be a useful way to provide this information, or detailed examples of successful approaches that can be copied.

A different set of issues are needed when teachers know what needs to be done to improve teaching and learning but lack authority or resources to implement these changes. In these circumstances, evaluation needs to provide a credible argument about the need for these resources or authority and fulfil an advocacy role.

Sometimes, however, the problem is not a lack of knowledge about how to improve, nor the resources to do this, but the willingness to make changes. This can be the case when teachers would be reluctant to undertake a major revision of a course and teaching material to address issues – since this would take time away from other activities such as

scholarly publication, which is prioritised in decisions about tenure and promotion. In such cases, what is needed is not just an evaluation but an evaluation system which addresses the incentives that encourage people to take particular actions. Evaluation in such circumstances needs to provide clear signals (in the form of ‘carrots’ – positive incentives – and ‘sticks’ – negative sanctions) about what is truly valued in the organisation.

Finally, an evaluation may be primarily intended to help teachers and students to develop their capacity for ongoing learning. While the initial findings might be small in scale, the real value lies in laying the foundation for ongoing learning about teaching and learning, and about ways to improve it.

Some Particular Evaluation Approaches

With these different types of evaluation in mind, the next section of the paper discusses briefly some diverse approaches to evaluating university teaching and learning:

- Performance indicators
- Student assessment, including authentic assessment and high stakes assessment
- Evidence-based policy and practice
- Peer review
- Appreciative inquiry

Performance Indicators

Performance monitoring works through providing information to check implementation during the program. Performance measures and indicators can be developed from the program theory and/or by using established benchmarks.

These performance indicators can refer to inputs (for example class sizes, access to computers), to processes (for example, the time spent in particular modes of teaching and learning or on particular topics), and to short-term and long-term outcomes (such as completion rates, pass rates and employment rates).

Performance indicators are often used for management purposes, and are intended to support improvements in teaching and learning by identifying poorly performing courses (in terms of low demand, poor

student completions and low employment rates) and either improving them or terminating them.

The timeliness of performance information can cause problems – if there are delays in collecting or reporting data, then corrective action can over-correct (as shown in Australia by cyclic shortages and surfeits of teachers). Aggregate data can hide pockets of excellence or problems. Any single indicator can be misleading. As part of the repertoire, performance measures are an important component – by themselves they are no substitute for more comprehensive evaluation.

Student Assessment Including High Stakes Testing and Authentic Assessment

Since improved student learning is usually the intended outcomes of improved teaching, it is not surprising that student assessment is often used as one of the sources of evidence of outcomes.

There are risks, however, in using student assessment for evaluation purposes, particularly when the assessments focus on a narrow range of knowledge and skills – in particular encouraging ‘teaching to the test’ where teaching focuses only on content and skills that will be tested, diminishing performance in terms of broader skills and genuine engagement in learning. Concern about the effects of using “high stakes” tests for evaluation purposes have led the American Evaluation Association to issue the following statement:

Recent years have seen an increased reliance on high stakes testing (the use of tests to make critical decisions about students, teachers, and schools) without full validation throughout the United States. The rationale for increased uses of testing is often based on a need for solid information to help policy makers shape policies and practices to insure the academic success of all students. Our reading of the accumulated evidence over the past two decades indicates that high stakes testing does not lead to better educational policies and practices. There is evidence that such testing often leads to educationally unjust consequences and unsound practices, even though it occasionally upgrades teaching and learning conditions in some classrooms and schools. The consequences that concern us most are increased drop out rates, teacher and administrator deprofessionalization, loss of curricular integrity, increased cultural insensitivity, and disproportionate allocation of educational resources into testing

programs and not into hiring qualified teachers and providing sound educational programs. [i] The deleterious effects of high stakes testing need further study, but the evidence of injury is compelling enough that AEA does not support continuation of the practice.

(American Evaluation Association, 2005)

Improving the quality of assessment is clearly important. ‘Authentic assessment’ is one response to this imperative.

Table 3: Comparing Authentic Assessment and Traditional Assessment (from Wiggins, 1990a, as cited by Herrington and Herrington, 1998)

Authentic assessment	Traditional assessment
Direct examination of student performance on worthy intellectual tasks	Relies on indirect or proxy items
Requires students to be effective performers with acquired knowledge	Reveals only whether students can recognise, recall or ‘plug in’ what was learned out of context
Presents the student with a full array of tasks	Conventional tests are usually limited to pencil and paper, one-answer questions
Attends to whether the student can craft polished, thorough and justifiable answers, performances or products	Conventional tests typically only ask the student to select or write correct responses – irrespective of reasons
Achieves validity and reliability by emphasising and standardising the appropriate criteria for scoring varied products	Traditional testing standardises objective ‘items’ and the one ‘right’ answer for each
‘Test validity’ should depend in part upon whether the test simulates real-world ‘tests’ of ability	Test validity is determined by matching items to curriculum content
Involves ill-structured challenges that help students rehearse for the complex ambiguities of professional life	Traditional tests are more like drills, assessing static and toooftne arbitrary elements of those activities

In a similar way, assessment of university teachers has moved to more authentic assessment through the use of teaching portfolios. In addition to providing more credible evidence of teacher performance, the process of developing a teaching portfolio itself supports a more systematic and reflective approach to teaching, as well as a more collaborative one (Way, 2002).

Evidence-based Policy and Practice

While the term ‘evidence-based practice’ is fairly recent, its roots go back at least to 1921 and Taylor’s scientific management approach. Taylor focused on identifying the best way to do things, and on helping others to use this information either through documenting it clearly or by redesigning the task. For example, having found the optimum weight of material that could be shovelled, he designed a shovel that could not carry more.

More recently there have been extensive efforts internationally to identify, document, and disseminate descriptions of best practice or ‘What works’ in order to improve practice. Five different types of evidence-based practice can be distinguished, as shown in the following table.

These are very different ways of building evidence about ‘what works’ or ‘what works for whom’. The first three focus on evaluating courses, subjects, curriculum or teaching approaches and then using this

Table 4: Approaches to Evidence-based Practice (from Rogers and Williams, 2006)

Meta-analysis	Previous evaluations that meet particular methodological requirements are quantitatively synthesized to produce knowledge about ‘what works’ in terms of implementation approaches (e.g., the Cochrane Collaboration for evidence-based medicine, and the similar Campbell Collaboration for human services such as education, criminal justice)
Realist synthesis	Previous evaluations using a range of methods and designs are synthesized analytically to produce knowledge about generative mechanisms that can be triggered by practice within particular contexts (i.e., what underlying assumptions works for whom in which programs in what contexts?)
Proven practice	A specific implementation package is developed, rigorously evaluated, and then replicated with sufficient fidelity – a franchise-like approach to improvement
Corporate memory	Information about previous practice within the organization is used to inform current practice. This could include organizational myths, shared histories, and performance data. Knowledge management techniques are used to gather, record, and access this information
Reflective practice	Practitioners use information about their current practice and its results, using techniques such as performance monitoring, learning logs, portfolios, and individual or group reflection

evidence to inform decisions made elsewhere. The last two focus on using local evidence to guide action and decisions.

Peer Review

Under this label are two rather different approaches to evaluating university teaching and learning. ‘Peer review’ can be undertaken by known colleagues – usually consisting of pairs or small groups of teachers reviewing each others’ teaching. The process can be useful as a way of providing advice from a supportive colleague. It can also help increase the motivation for changing practice. Peer review requires careful negotiation of the scope of the review, including the period of time that will be spent, and its focus which might include:

- Review of course materials
- Review of student achievement and marking standards
- Feedback from tutors and demonstrators
- Observation of classes – using a guide or checklist
- Video recording of class presentation
- Sustained collaboration – some models in practice
- Structured discussion before or after classes
- Structured discussion of course design

‘Peer review’ involves recognised experts reviewing teaching practice and teaching infrastructure. It can lead to improved teaching and learning by providing specialist advice, and also increase the motivation to improve performance since it is often part of formal accreditation processes.

Appreciative Inquiry

Appreciative Inquiry (Elliott, 1999) is a fundamentally different approach to evaluation to others which often focus on identifying and correcting gaps and problems. Appreciative Inquiry is built on the assumption that people are more motivated by focusing on achieving good outcomes than on avoiding or fixing problems. It assumes the energy for organizational and program improvement is in focusing people on their aspirations and dreams of what could be. This means replacing the traditional problem solving approach – as reflected in the question ‘What is going wrong and how do we fix it?’ – by two simple questions: “What do we do well?”, and “How can we do more of it”?

An Appreciative Inquiry will essentially pass through four stages:

- Discovering periods and points of excellence within the program
- Dreaming an ideal program
- Designing new structures and processes
- Delivering the dream

When well-done, Appreciative Inquiry can help to gather the energy for change, as well as the courage to address known problems; when done poorly it risks a self-serving focus on positive aspects which leads to uncritical and incomplete assessments of program performance.

Conclusion

This paper has outlined a range of different types of evaluation and approaches to undertaking evaluation. Considering these different ways in which evaluation is intended to improve university teaching and learning, as well as the different types and approaches, will help to develop the right sort of evaluation for a given situation.

References

- Åkerlind, G., and Pettigrove, M. (1996). *Peer review and self-evaluation: A toolkit resource booklet for the improvement of teaching and learning*. Canberra: Australian National University.
- Batchelor, S., and Norrish, P. (2005). *Framework for the assessment of ICT pilot projects: Beyond monitoring and evaluation to applied research*. Washington: infoDev Program, The World Bank.
- Bradshaw, J. (1972). A taxonomy of social need, *New Society* (March) 640-3.
- Duignan, P. (2004). Intervention logic: How to build outcomes hierarchy diagrams using the OH Diagramming Approach [WWW document]. From The StrategicEvaluationWebSite.URL
<http://www.strategicevaluation.info/se/documents/124pdff.html>
- Dunkin, M. (1997). Assessing teachers' effectiveness'. *Issues in Educational Research*, 7(1), 37-51.

- Elliott, Charles (1999). *Locating the energy for change: An introduction to appreciative inquiry*. Winnipeg: International Institute for Sustainable Development. Available for download at <http://iisd1.iisd.ca/pdf/appreciativeinquiry.pdf>
- Funnell, S. (2000). Developing and Using A Program Theory Matrix for Program Evaluation and Performance Monitoring in Rogers, P. J., Petrosino, A. J., Hacsı, T. and Huebner, T. A. (eds). (2000) *Program theory evaluation: Challenges and opportunities*. (New Directions in Evaluation series). San Francisco: Jossey-Bass.
- Guice, R. & da Costa, C. & Rogers, P.J. (1996). Quantitative methodology: An integrated multimedia approach to teaching one-way anova in Nursing Research - Project Management and Design and Development Issues, *Evaluation News and Comment*, 5(2), p. 57.
- Hadley, J., Wall, D. and Khan, K. (2007). 'Learning needs analysis to guide teaching evidence-based medicine: knowledge and beliefs amongst trainees from various specialities'. *BMC Medical Education*, 7: 11.
- Herrington, J. & Herrington, A. (1998). Authentic assessment and multimedia: How university students respond to a model of authentic assessment. *Higher Education Research and Development*, 17(3), 305-322.
- Ingvanson, L. and Rowe, K. (2007). *Conceptualising and Evaluating Teacher Quality: Substantive and methodological issues*. Paper presented at the Economics of Teacher Quality conference Australian National University, February 2007.
- Kippen, C. (1995). *Planning a degree programme in podiatry, based on a medical problem solving curriculum and developed from an industrial needs analysis*. Paper presented at the 4th Annual Teaching Learning Forum, Edith Cowan University, February 1995.
- Lee, F. and Yeap, B.H. (2005). The use of educational technologies in university teaching & learning. Paper presented at the third International Conference on Multimedia and Information and Communication Technologies in Education, Cáceres (Spain), in A.

Méndez-Vilas, B. González-Pereira, J. Mesa González, J.A. Mesa González (eds) *Recent research developments in learning technologies*. FORMATEX, Badajoz, Spain.

Malcolm G. Pettigrove (2008). American Evaluation Association, Position Statement on High Stakes Testing In PreK-12 Education. Available at <http://www.eval.org/hst3.htm>

NSLP (National Service Learning Partnership) (2006). 'Logic models'. Available at <http://nslp.convio.net/site/DocServer/Tool2LogicModels.pdf?docID=2282>

Owen, J. (1992/2007). *Program evaluation: Forms and approaches*. (1st and 3rd editions) Allen & Unwin/Sage Publications.

Rogers, P.J. (2000). Program theory evaluation: Not whether programs work but how they work, in G. Madaus, D. Stufflebeam & T. Kellaghan, (eds) *Evaluation Models*, New York: Kluwer Press.

Rogers, P.J. and Williams, B. (2006). Evaluation for practice improvement and organizational learning in I. Shaw, J. Greene and M. Mark (eds) *Sage Handbook of Evaluation*, London: Sage Publications.

Rossi, Peter H., Freeman, Howard H. and Lipsey, Mark W. (1998/2003). *Evaluation, A Systematic Approach* (6th edition/7th edition), Sage Publications.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné & M. Scriven (Eds.), *Perspectives of curriculum evaluation*, 39-83. Chicago, IL: Rand McNally.

Scriven, M. (1980). *The Logic of evaluation*. Inverness: Edgepress.

Scriven, M. (1990). Teacher selection. In J. Millman & L. Darling-Hammond, *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers*, (pp. 76-103). Newbury Park, CA: Sage.

- Scriven, M. (1991). *Evaluation thesaurus*. (4th ed.) Newbury Park, CA: Sage Publications.
- Suchman, E. (1967). *Evaluative research*. New York: Russell Sage Foundation.
- Wagner, Daniel A., Bob Day, Tina James, Robert B. Kozma, Jonathan Miller and Tim Unwin. (2005). *Monitoring and evaluation of ICT in education projects: A handbook for developing countries*. Washington, DC: infoDev /World Bank. Available at: <http://www.infodiv.org/en/Publication.9.html>
- Way, D. (2002). Evaluating teaching portfolios: Moving from a competitive to a collaborative culture. Paper presented at the *4th World Conference of the International Consortium for Educational Development in Higher Education*, Perth, Australia <http://www.csd.uwa.edu.au/iced2002/publication/index.html>.
- Wiggins, G. (1990a). *The case for authentic assessment*. Washington, DC: ERIC Clearinghouse on Tests, Measurement, and Evaluation. (ERIC Document Reproduction Service No. ED 328 606).
- Wiggins, G. (1990b). *The case for authentic assessment. ERIC digests*. Available at <http://ericae.net/edo/ED328611.htm>
- W.K. Kellogg Foundation (2004). *Logic model development guide*. Battle Creek, MI: KelloggFoundation. Available at <http://www.wkkf.org/Pubs/Tools/Evaluation/Pub3669.pdf>