Systematically evaluating the effectiveness of quality assurance programmes in leading to improvements in institutional performance

Deirdre Lillis

Routledge
Taylor & Francis Group

Systematically evaluating the effectiveness of quality assurance programmes in leading to improvements in institutional performance

DEIRDRE LILLIS*

Dublin Institute of Technology, Dublin, Ireland

Higher education institutions worldwide invest significant resources in their quality assurance systems. Little empirical evidence exists that demonstrates the effectiveness (or otherwise) of these systems. Methodological approaches for determining effectiveness are also underdeveloped. Self-study-with-peer-review is a widely used model for ensuring the quality of the core teaching, research and engagement activities of higher education institutions. This article illustrates how an established social-programme evaluation methodology can be used to determine its effectiveness in leading to improvements in institutional performance. The concept of effectiveness and the particular challenges posed by the higher education organisational culture are considered. An example of the systematic evaluation of three self-study programmes is provided to illustrate the concept. It is concluded that social-programme evaluation has significant potential in evaluating the effectiveness of quality assurance initiatives in higher education.

Keywords: evaluation of effectiveness; social-programme evaluation; quality assurance; higher education; self-study-with-peer-review

Quality assurance in higher education

National quality assurance agencies, almost unheard of 20 years ago, are now in place in almost all Organisation for Economic Cooperation (OECD, 2003) countries. Stensaker (2006) noted that while there are a growing number of studies on quality assurance, there is a lack of research on the impact of quality assurance at institutional level. He cited the methodological issues surrounding the assessment of the impact of quality assurance processes as a major challenge. Harvey and Newton (2004) noted that establishing definitive causal links and isolating their effects from other factors is a difficult task. Birnbaum (2000, p. 10) stated that there are 'few published examples

*Email: deirdre.lillis@dit.ie

in the academic sector of attempts to assess the institutional consequences of a management fad through data that provide evidence either of organisational outcomes or of the satisfaction of users'.

Evidence of the effectiveness of the core activities of higher education (teaching, research and engagement) is generally increasing (Pascarella & Terenzini, 2005). It is important to note that this article does not focus on the effectiveness of these core activities but rather the effectiveness (or otherwise) of the quality assurance instruments that are used to assess them. These instruments are in widespread use by governments, higher education quality assurance agencies and internally within institutions. The question being addressed is to what extent these instruments can be trusted.

Van Vught and Westerheijden (1995) found that the predominant model for quality improvement includes regular self-evaluations with external peer review by the higher education institute. Self-study-with-peer-review is often cited as being most suited to the 'professional bureaucracy' type of organisation (Mintzberg & Quinn, 1998) as it gives ownership for quality to the institution concerned. Kells (1992) noted that the external driver for self-study programmes usually relates to accreditation status but that self-study programmes often have additional internal aims. As a form of quality assurance in higher education, self-study programmes can take context into account, can straddle academic disciplines and are generally accepted by the academic community. Limitations of the model include its significant overhead, the length of time needed to complete a full cycle and the necessity of taking staff away from their core duties. Therefore, it is reasonable to ask what the return on this investment is, whether the model is fit for purpose and what substantive, additional improvements have resulted from the self-study-with-peer-review exercise. Massey (2003) highlighted that public trust in higher education is being eroded and being able to demonstrate the reliability and validity of quality assurance instruments to external stakeholders is essential for continued credibility.

El-Khawas (1998) noted that most policy research has focused on institutional-level effects even though the impact of self-study-with-peer-review programmes often depends on the reaction of departments and individuals. Sallinen *et al.* (1994) noted institutional impacts of self-study-with-peer-review that included improving transparency, communication, organisational learning, effectiveness and readiness for change. Henkel (2004, p. 27) argued that self-study exercises could create 'new levels of understanding and mutual interest in a department'. Using a systematic evaluation methodology, Thorn (2003) found that self-study-with-peer-review led to an increased awareness of strategic planning, gave staff a forum for input to decision-making and noted the failure in some instances to face up to weaknesses. Notwithstanding the above studies, and despite its widespread use in higher education, there is a significant lack of empirical research that demonstrates the effectiveness of self-study-with-peer-review.

What is meant by effectiveness?

Social-programme evaluation is widely used in the public and non-profit sector for undertaking research into the effectiveness and efficiency of programmes (Patton, 2002; Rossi *et al.*, 2004). It has applications in domains where planned interventions are made to bring about improvements in people's lives (for example, healthcare, social care, environment and public sectors). The social nature of programmes in these sectors share many similarities with quality assurance initiatives in higher education in their complexity and in the inherent difficulties of isolating the net effects of the programme from what would have happened anyway. Such evaluations are challenging and it is not possible to give definitive answers about their net impact but it is still possible to aim to give a credible estimate of the impact of a programme or intervention. While alternative evaluation approaches exist, social-programme evaluation was chosen for two reasons. First, it is accessible to a 'lay' reader while losing none of its rigour. Second, its relatively widespread use in many different fields allows for comparability between studies.

The first concept to be explored is what is meant by effectiveness. The classic interpretation is that an effective programme is one that meets its stated goals and objectives. This leads to a rational, 'goals-oriented' evaluation approach (Vedung, 1997). Problems can occur when goals are poorly articulated or not prioritised and it is possible that unanticipated side effects, both good and bad, are ignored. A broader concept of effectiveness is therefore required, which also allows for improvements arising from programmes over and above what was intended or stated in the goals and objectives. This is termed 'prescriptive valuing' and leads to a 'goals-free' approach (Van der Knaap, 1995). To give a rounded and credible estimate of the impact of a programme, effectiveness can be defined as: (i) a programme must meet its stated goals and objectives; and (ii) it may lead to additional (possibly unintended) improvements.

Going beyond subjective opinion

A reality-oriented post-positivist standpoint underpins this approach. Results can be viewed as probable causal effects and the reader has discretion to draw his or her own conclusions on the basis of the evidence presented. The goal is to minimise subjectivity and provide objective evidence of actual programme impact. The fundamental hypothesis tested is that 'the programme is effective in leading to improvements in performance'. Birnbaum (2000) noted that the private sector typically seeks empirical data to evaluate major management innovations whereas the higher education sector relies primarily on subjective judgement. Social-programme evaluation seeks to find empirical evidence of effectiveness beyond the views of those involved in the programme. In social-programme evaluation, the 'judgments

of experts, programme administrators and participants' who are asked to make 'assessments of how the programme has affected them' are used sparingly and with caution (Rossi *et al.*, 2004, p. 217). The main concern is when such judgements are used definitively in isolation from other sources of data such as the document record of the institution. This emphasis on objectivity and triangulation of data sources has the potential to strengthen many studies on quality assurance in higher education. This in turn goes some way to addressing stakeholder perceptions of credibility.

How is the impact of a programme evaluated?

Using a social-programme evaluation approach, a programme is systematically evaluated using a four-step process as follows (Rossi *et al.*, 2004):

- assessment of the need for the programme;
- assessment of the process design;
- assessment of the impact of the programme;
- separating net from gross outcomes.

Assessment of need for the programme

There is always a danger in any organisation that things are done the way they have always done them, blindly following convention without critical questioning of the need to do something. Rossi *et al.* (2004) noted that evaluation of established programmes rarely focuses on the underlying conceptualisation as stakeholders are often reluctant to question tradition unless prompted to by exceptional circumstances. Self-study-with-peer-review is a widely accepted method of quality assurance in higher education. Without empirical evidence to support its effectiveness, an assessment of the need for the self-study programmes is considered important. Such an assessment clarifies the goals of the programme and considers alternative approaches. This questioning is required at all levels within the sectors, both within institutions and within the agencies that require institutions to undertake regular self-study programmes.

Assessment of process design

The design of a programme is important for two reasons. First, a programme may be badly designed, making it unfit for purpose and unable to achieve the intended outcomes (for example, a template for a self-evaluation report may not contain the appropriate headings). Second, it is possible that a well-designed programme may be badly implemented (for example, inadequate attention is paid to the selection of panel members with appropriate expertise, leading to inappropriate recommendations).

An assessment of process design also determines the extent to which the programme theory 'as-intended' was actually implemented. It is difficult to accurately assess the impact of programmes that have been partially or incorrectly implemented. For the self-study-with-peer-review model the process assessment concentrates on the main activities as follows: (i) internal self-evaluation of activities; (ii) self-study report; (iii) peer-review process; (iv) implementation of peer-review recommendations and other improvements identified.

Patton (2002) noted that evaluations place varying degrees of emphasis on programme process evaluation. Peer-review panels often explore the process undertaken for internal self-study in as much detail as the actual outcomes, as an indication of how valid the conclusions are. Undertaking a process evaluation also increases the generalisability of the research by providing a documented frame of reference for future evaluations. It also helps to distinguish between 'espoused theory' (what those involved would like to think happened) from the 'theory-in-action' (what actually happened) (Patton, 2002).

Assessment of the impact of the programme

The purpose of clarifying programme impact theory is to determine in what way programme activities effect changes. It is generally illustrated in a logic diagram (Figure 1) and is developed from the perspective of capturing the programme 'as-intended'.
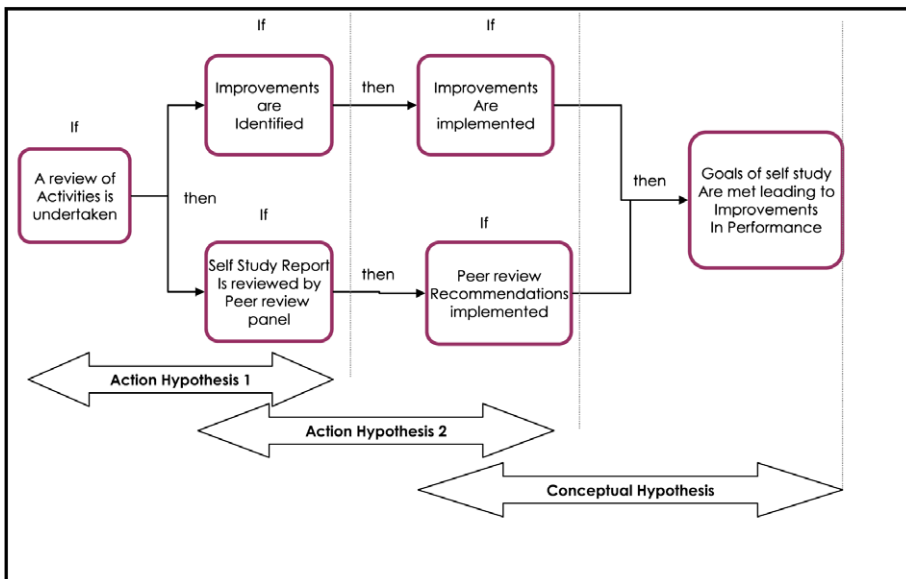

Figure 1.   Programme impact theory (self-study-with-peer-review).

Detailed process descriptions can be reconstructed and fully documented from the document record or from interviews with participants. Impact theory is based on the contention that outcomes that are a direct result of the programme (proximal outcomes) must be evaluated if longer-term outcomes (distal outcomes) are to lead to improvements (Rossi *et al*., 2004). In other words, the attainment of the overall goal of the programme is dependent on the attainment of intermediate outcomes such as the implementation of improvements identified by the internal team or by the expert panel. By way of example, updates to course syllabi (a proximal outcome) lead to improvements in the relevance and quality of the course (a more distal outcome). While some of the richness of the programme may be lost in this approach, it is necessary to break down its complexity into a model that lends itself toward measurement.

As discussed earlier, to be considered effective, programmes must meet their goals and objectives, requiring a goals-based impact assessment. There is also provision for programmes leading to other (possibly unintended) improvements, requiring a goals-free impact assessment.

Goals-based impact assessment

The classic goals-based impact assessment is used to evaluate the extent to which programmes meet their stated goals and objectives. Rossi *et al*. contended that the ideal impact assessment design, if somewhat unrealistic for complex social programmes, is an experimental design (Rossi *et al*., 2004). This assumes that programmes are stable processes with pre-determined outcomes that can be represented by independent variables in a quantifiable fashion and where relationships between variables can be portrayed statistically.

The complexity and relative instability of social programmes means that a full experimental design is not possible and that a quasi-experimental design must be used instead. A reflexive 'time series analysis' design is the strongest of these approaches (Rossi *et al*., 2004). This captures the impact of the programmes at three or more points in time including (i) the period before the programme started, (ii) a mid-way point and (iii) the period after the programme. For example, in the case of the self-study programmes, the period before the self-study began, the panel visit and one year after the panel visit are the time intervals that could be used. The selection of these time points is context-specific and needs due consideration as these may have a bearing on the outcomes. At these time points, objective evidence of completion of programme objectives is sought from various sources (for example, an acknowledgement by the external peer-review panel or the proceedings of relevant fora in the institution, such as the academic council, senior management team or governing body). These can be supplemented by interviewing informants when necessary but it is essential to use triangulated data sources to minimise the reliance on subjective opinion.

When the goals of a programme are complex a key question that arises is: what percentage of the goal must be complete for the overall goal to be considered complete? For example, if 80% of the recommendations of the external peer-review panel have been implemented, can it be said that the goal of the programme has been met? The threshold set for the 'percentage complete' is therefore a key consideration and the determination of this threshold is not a straightforward exercise. The views of programme stakeholders, the literature base available for comparative purposes and the specific context of the programme are all important factors.

The programme impact theory states that outcomes that are a direct result of the programme (proximal outcomes) must be evaluated if longer-term outcomes (distal outcomes) are to lead to improvements; that is, the goals of the self-study are dependent on the implementation of the improvements identified and the external peer-review recommendations. The aim of the goals-based element of the impact assessment is, therefore, to provide a credible estimate of the impact of the programmes. It must be acknowledged that this is not the ideal approach, that the resulting estimates of programme impact are not definitive and that the potential of bias must be actively counteracted. The quasi-experimental approach is nevertheless a feasible approach to take.

Goals-free impact assessment

House (1993) argued that goals-free evaluations are very challenging to do and Scriven (1972) suggested that a goals-free evaluation should run in parallel with a goals-based evaluation for maximum effect. This ensures that the dynamic nature of the programmes can be accommodated even with the complexity of a changing environment (Patton, 2002). The aim is to capture improvements that may have resulted from the programme but that were not explicitly stated in the goals and objectives (Patton, 2002). Although these improvements may not have been anticipated it does not mean that they are not important and they can have positive or negative impacts (Rossi *et al.*, 2004). For example, the self-study programmes may have led to improvements in building shared vision among staff, enhancing the leadership capabilities of the management team or clarifying future direction. None of these were explicitly stated but are arguably as important an outcome as the stated objectives of the programme. Rossi *et al.* (2004) noted that the first-hand accounts of programme informants is a good source of information for these types of impacts.

Separating net from gross outcomes

The difficulties of isolating the impacts of quality assurance programmes from other factors is highlighted in the literature (Harvey & Newton, 2004).

The social nature of the programmes, the complexity of the environment and the number and range of participants makes it almost impossible to make definitive or positivistic statements in relation to this. This is not unique to higher education and it is a common problem for most complex social programmes. Separating net from gross outcomes is the most problematic but most critical aspect of programme evaluation, which entails identifying what happened as a result of the programme compared to what would have happened anyway. Rossi *et al*. (2004) noted that the estimation of true programme impact is the most demanding evaluation research task. Therefore, results must be presented in probable terms. In essence this attempts to answer the questions 'what would have happened anyway?' For example, it is almost certain that an institute or faculty would respond to

Table 1.   Taxonomy of result types (Pollitt & Bouckaert, 2004).

Result type	Level	Description
Operational	1	Discrete and quantifiable results or efficiency measures. Examples include: objectives with targets relating to student numbers; retention rates; specific resources or facilities; targeted marketing initiatives; development of new courses; implementation of specific initiatives (e.g., school visits programme).
Process	2	Improved management or decision-making processes that are linked directly to actual improvements. Examples include: developing links with stakeholders for a specified purpose (e.g., teacher training, assisting schools with specific projects); introducing change to organisational structure (e.g., new position created for specific purpose); developing and implementing a strategy or plan for a specific functional area (e.g., develop a marketing plan; encourage or facilitate staff to participate in research or consultancy; ensure an equitable workload for students; investigate new markets or new areas; investigate a new course-development strategy).
Capacity	3	Systems-level outcomes that enhance the capacity of the organisation. Examples include: continued development of some activity without specified outcomes (e.g., developing links and partnerships, improving quality or overall student experience, encouraging teaching excellence, encouraging campus company startups). Change in organisational culture (e.g., managing in more open and consistent manner, managing in a more effective and efficient manner); development of centre of excellence; contribute to national policy.
Ideological	4	Movement of organisation toward desired or ideal state. These are intangible but desirable states (e.g., total quality culture in all operations, foster an entrepreneurial ethos, enhance standing as a contributor to regional development).

changes in its environment in various ways, irrespective of ever undertaking a self-study programme. Improvements that found their origin in the normal day-to-day activities of the institution must be systematically identified and tracked in the document record and then separated from improvements that found their origin in the programmes. Through thorough document analysis, each issue that arises during the time series is tracked from when it first appeared in the document record to its eventual completion, retirement or abandonment.

Particular challenges for higher education

The ease by which programme goals can be measured and evaluated is a key concern. There is a strong argument in the higher education literature that it is impossible to define any single combination of performance indicators that appropriately measure performance (Kells, 1992; Linke, 1992). Higher education is not unique in this regard, however, as many social programmes face similar challenges.

It is much easier to accurately assess the impact of programme when a high percentage of its objectives are written in measurable terms. Poorly articulated goals such as 'produce good quality research' are ambiguous and difficult to measure, whereas 'have 10 papers published in peer-reviewed journals' is more easily evaluated (noting that this is not necessarily an adequate indicator of quality either).

In the wider public-sector management literature, Pollitt and Bouckaert (2004) provided a mechanism by which the type of result from a programme can be categorised on the basis of the extent to which the result is evaluable or measurable. Results are categorised as being operational, process, capacity or ideological. This is outlined in Table 1, with examples adapted for higher education.

Operational results are typically expressed quantitatively and compared with some preset standard (for example, this year's student intake compared to last year). Process results are expressed as the effect of improving activities (for example, increasing graduate throughput while maintaining the quantity and quality of the student intake might suggest that the teaching process has improved). However, process results need to be coupled with quality and cost data (for example, academic standards may have dropped to ensure a constant throughput of graduates). Capacity-level results are improvements in either structures or culture, leading to organisations that are more flexible, that have a higher capacity to learn and are more responsive.

In a programme evaluation, this classification is used to provide a global assessment of how evaluable are the goals and objectives of the programme. For example, if a self-study programme has only a small percentage of its objectives at operational or process level it will be difficult to glean mean-

ingful insights into its impact. Further work may be necessary to translate goals written in capacity or ideological terms into more measurable goals that lend themselves to evaluation.

The systematic evaluation of the effectiveness of three self-study-with-peer-review programmes

The following is a worked example to illustrate the application of the social-programme evaluation methodology described in this article. The institutional impacts of three self-study programmes undertaken during an eight-year timeframe (1997–2006) in one Irish Institute of Technology are evaluated. The Institute was required to undertake quinquennial institutional and school (faculty-level) reviews that entailed comprehensive self-studies with external peer review. The first self-study programme was called 'programmatic review' (PR1) and was essentially a school review (including a review of all teaching and research courses). The second self-study programme was called 'delegated authority' (DA1) and was at institutional level for the purposes of gaining degree-awarding authority. The third self-study programme was a second programmatic review (PR2) in the same school five years later. The external peer-review panels commended the thoroughness of all three self-study programmes, indicating that they are likely to provide good examples and will provide an information-rich case study. This meets the criteria of an 'intensity case', which is a case that is not unusual but from which much can be learned (Patton, 2002).

The study straddles an eight-year time period and the main data sources used were institute documents including the reports of the self-study programmes, proceedings of the governing body, the academic council, the senior management team, school boards and course boards. Semi-structured interviews were conducted with 17 key informants who had a major involvement with the programmes. These included all of the senior management team and approximately half of the heads of academic departments and central services managers.

An assessment of the need for the self-study-with-peer-review programmes

Although the driving force for all three programmes was ultimately to meet external requirements linked to the accreditation status of courses of study, they were also seen as opportunities to progress internal objectives. Given the scope of the three programmes, it is likely that meeting the external requirements would by default bring many internal improvements also. The goal of the delegated authority programme (DA1) was essentially to achieve self-awarding status following an institute-wide review of all activities but four additional internal objectives were also set. These included the implementation of a strategic management and continuous improvement

framework. The goal of the programmatic reviews was to ensure '(a) quality improvements are made to programmes of higher education and training and (b) programmes remain relevant to learner needs, including academic and labour market needs'. There were internal objectives also that related to specific objectives from the institute's strategic plan for implementation in the school (including modularisation of courses and the development of flexible modes of delivery). In essence all three programmes were needed as they were required by the quality assurance system.

Assessment of the self-study-with-peer-review process

An assessment of process design was undertaken to determine the extent to which the programme theory 'as-intended' was actually implemented, as it is difficult to assess the impact of programmes that have been partially or incorrectly implemented. In summary, all components of all three self-study programmes were completed largely 'as-intended' as evidenced by the documents associated with each phase (for example, self-study report, panel report).

An assessment of the impact of the self-study-with-peer-review programmes

Three time points for each of the programmes were set to capture progress before, during and after each programme. Evidence of completion of the goals and objectives was sought, using the document record primarily (for example, minutes of meetings, progress reports). An element of subjective judgement is unavoidable in setting the threshold for 'percentage complete' and for this reason three possible threshold values are illustrated to allow readers to draw their own conclusions.

At the 50% threshold value, all objectives of all the self-study programmes were met (Table 2). The only deviation is at the 66% threshold for PR1 and PR2 (noting the shorter timeframe for the improvements from PR2 to be implemented).

DA1 was in essence a summative evaluation in that it made a judgement as to whether or not the institute met the criteria for delegated authority. No specific recommendations for improvement were made by the external review panel. Three-quarters (75%) of the peer-review recommendations for

Table 2.   Meeting stated objectives.

Threshold	DA1	PR1	PR2
⩾ 33%	4 of 4 (100%)	7 of 7 (100%)	10 of 10 (100%)
⩾ 50%	4 of 4 (100%)	7 of 7 (100%)	10 of 10 (100%)
⩾ 66%	4 of 4 (100%)	5 of 7 (71%)	9 of 10 (90%)

Table 3. Meeting goals and objectives.

Ref	DA1	PR1	PR2
Stated objectives and peer-review recommendations	4	19	23
Objectives completed	4 of 4 (100%)	16 of 19 (84%)	14 of 23 (61%)

PR1 were completed and 30% for PR2 were implemented (the shorter time-frame of the impact assessment should be taken into account when interpreting the PR2 results). Almost all recommendations made to courses of study were implemented within a short time period after the review. Revisions to courses took effect for the next intake of students to the courses.

The programme impact theory states that the proximal outcomes for the self-study programmes are the objectives of the self-study including implementing the peer-review panel recommendations. At the 50% threshold, DA1 met 100% of its objectives, PR1 met 84% and PR2 met 61% (noting the shorter timeframe for PR2) (Table 3).

Arguably, therefore, in gross terms, the self-studies were effective in leading to improvements.

Other improvements arising from the self-study programmes were identified by asking informants what positive and negative impacts the self-study programmes had. Nearly half the informants ($n = 7$) began their answer by stating that they did not see any negatives with the self-study process. The positive impacts most frequently cited by informants included concepts such as: the overhead involved ($n = 9$), building commitment ($n = 8$), the opportunity to review activities ($n = 6$) and involve stakeholders ($n = 3$). Informants were also asked 'Can you think of an example of something which wouldn't have happened without the self-study process?'. As expected, many informants stated that it was a difficult question to answer or took more time before answering the question. Notwithstanding this, over two-thirds of the informants ($n = 13$) could think of a specific example of something they felt would not have happened without the self-study process. These included ideas for new course development ($n = 4$), documentation of quality assurance procedures ($n = 3$), prioritisation of research ($n = 2$) and cross-departmental team working ($n = 2$). Two informants that could not think of a specific example still thought that certain things would not have happened without the self-study process. One stated that without self-study 'everything just stagnates, there's no fresh thinking'.

Separating net from gross outcomes

The outcomes of the three self-study programmes were categorised as either originating within the programme or as outside it (Table 4).

Table 4. Summary of origin of objectives including peer review recommendations of self-study programmes.

	DA1	PR1	PR2
Total objectives	4	19	23
Completed/ongoing objectives originating within the programme	3 of 4 (75%)	7 of 19 (37%)	7 of 23 (30%)

By way of example, one objective of the programmatic review pro-gramme (PR1) was 'To review the development of the courses over the pre-vious five years with particular regard to the achievement and improvement of quality'. The PR1 process was the only mechanism within the quality assurance system by which substantive changes to courses of study could be made and therefore it can be clearly stated that this objective would not have been achieved without PR1. On the other hand, one of the objectives of PR2 was 'To review the plans (of the School) for future development'. At the time of PR2, each department had produced a strategic plan that was subject to an annual internal review and it is possible, therefore, that this objective could have happened without PR2.

Extensive document analysis was used to trace the origin of the objec-tives and they were analysed from the perspective of whether they would have happened regardless of the programmes. In summary, 75% of the com-pleted objectives of DA1, 37% of the completed objectives of PR1 and 30% of the completed objectives of PR2 can be ascribed to the programme (they would not have happened without the programme). In summary, at least a third of the net improvements would not have happened without the pro-grammes.

Summary of programme evaluation

It has been established that there was a need for the self-study programmes and that the programmes were implemented largely 'as-intended'. The pro-grammes were effective as the substantial majority of their objectives and peer-review recommendations were completed. Informants also perceived the programmes to be effective. Three quarters of the outcomes of DA1 and approximately one-third of the outcomes of PR1 and PR2 could be ascribed to the programme (net outcomes).

Lessons learned and wider implications

Much can be learned from the social-programme evaluation literature and it has significant potential as a robust and versatile methodology for systemati-cally evaluating the effectiveness of quality assurance processes in higher

education. It has the added advantage of being accessible to a lay readership and providing a framework that enables comparisons to be made across numerous case studies and across sectors. Key questions remain for quality assurance agencies and higher education institutes. The most fundamental are whether tried-and-trusted processes for quality assurance are effective in leading to improvements and how do we know? The importance of context in higher education research means that institutions and agencies will often have to answer these questions for themselves. The overhead involved in self-study programmes is significant and the question of whether the benefits outweigh the costs is an important one.

References

Birnbaum, R., 2000, 'The life cycle of academic management fads', *Journal of Higher Education*, 71(1), pp. 1–16.

El-Khawas, E., 1998, 'Strong state action but limited results: perspectives on university resistance', *European Journal of Education*, 33(3), pp. 317–20.

Harvey, L. & Newton, J., 2004, 'Transforming quality evaluation', *Quality in Higher Education*, 10(2), pp. 149–65.

Henkel, M., 2004, 'The demise of a dominant culture? Higher education institutions in transition', *Learning and Teaching in the Social Sciences*, 1(1), pp. 21–32.

House, E.R., 1993, *Professional Evaluation* (London, Sage Publications).

Kells, H.R., 1992, *Self-regulation in Higher Education* (London, Jessica Kingsley).

Linke, R.D., 1992, 'Some principles for the application of performance indicators in higher education', *Higher Education Management*, 4(2), pp. 194–203.

Massey, W.F., 2003, *Honouring the Trust: Quality and cost containment in higher education* (Bolton, MA, Anker Publishing).

Mintzberg, H. & Quinn, J., 1998, *Readings in the Strategy Process* (Englewood Cliffs, NJ, Prentice Hall).

Organisation for Economic Co-operation and Development (OECD), 2003, *Education Policy Analysis* (Paris, OECD).

Pascarella, E.T. & Terenzini, P., 2005, *How College Affects Students: A third decade of research* (San Francisco, CA, Jossey Bass).

Patton, M.Q., 2002, *Qualitative Research and Evaluation Methods*, 3rd edn. (London, Sage Publications).

Pollitt, C. & Bouckaert, G., 2004, *Public Management Reform* 2nd edn. (Oxford, UK, Oxford University Press).

Rossi, P., Lipsey, M. & Freeman, H., 2004, *Evaluation: A systematic approach* (London, Sage Publications).

Sallinen, A., Konttinen, R. & Panhelainen, M., 1994, 'Interactive model of self evaluation quality assessment at the University of Jyvaskyla: a pilot study', *Higher Education Management*, 6(3), pp. 357–75.

Scriven, M., 1972, 'Prose and cons about goal-free evaluation', *Evaluation Comment*, 3(1), pp. 1–7.

Stensaker, B., 2006, Impact of quality processes, paper presented at the 1st *European Forum for Quality Assurance*, Munich, Germany, 23–25 November.

Thorn, R., 2003, 'Self study in an Institute of Technology: an evaluation', paper presented at the *European Association of Institutional Researchers Forum 2003*, Limerick, Ireland, 24–27 August.

Van Vught, F. & Westerheijden, D.F., 1995, 'Quality measurement and quality assurance in European higher education', in Bouckaert, C. (Ed.), *Quality Improvement in European Public Services* (London, Sage Publications).

Van der Knaap, P., 1995, 'Policy evaluation and learning: feedback, enlightenment or argumentation', *Evaluation*, 1(2), pp. 189–216.