

The Evaluation of Educational Innovation

CECILIA JACOBS

Peninsula Technikon, South Africa

This article presents a model for the evaluation of educational innovation in a context of transformation. The model incorporates formative, summative and illuminative evaluation goals and emphasizes the need to locate the innovation which is being evaluated within the context and policy framework of its operation. The evaluation framework provided by the model takes into account the full range of variables impacting on innovative educational practice and subjects the innovation, along with its transforming educational context, to the scrutiny of evaluation. The article thus presents a 10-stage generalized evaluation model which can be used as a framework for the evaluation of any type of educational innovation. Each stage in this 10-stage evaluation process is expanded on, discussed in detail in the article and illustrated with examples from the evaluation of an innovative intervention at a tertiary institution in South Africa.

KEYWORDS: contextually oriented model; evaluative framework; institutional transformation; quality assurance; tertiary sector

Introduction

Globally, educational evaluation is increasingly becoming a part of educational policy. This is especially true in the United Kingdom. In South Africa, where policy is in the process of being formulated for a transformed education system, a keen awareness of the importance of evaluation and quality assurance in education is evident in state policy documents. The state-appointed National Commission on Higher Education (NCHE), in its final report to the Minister of Education on a new policy framework for higher education transformation, proposes a Higher Education Quality Committee (HEQC) as an umbrella body for co-ordinating quality assurance in higher education, with specialist bodies undertaking the external evaluation function (NCHE Report, 1996: 109). The report further proposes that such a system be underpinned by a combination of institutional self-evaluation and external evaluation. This awareness, especially regarding evaluation in the tertiary education sector, brings a welcome change to an education sector previously lacking in a culture of evaluation. However, since

the entire education system finds itself in a period of transformation and the needs of the various academic communities are constantly changing, educators are challenged to become innovative practitioners in their attempts to address these challenges. While academic communities in South Africa, for the most part, are aware of the need for individual and institutional transformation, there are a range of interpretations as to how and where transformation should take place, all of which is compounded by a lack of shared understandings regarding particular educational innovations. Most innovation therefore occurs outside of the funded, mainstream functioning of institutions. This context, within which innovation operates, requires an extensive evaluation framework, such as the one expounded in this article, which takes into account the full range of variables impacting on educational innovation.

The Model

The model was developed by the author after a process of literature review and consultation with various stakeholders participating in a particular educational innovation at Peninsula Technikon (a tertiary institution in South Africa), as well as the actual evaluation of the innovation. For the purposes of this article the term educational innovation will refer to any form or type of educational practice which is new to or only marginally implemented in a particular academic context, and which is designed to develop, improve, make more relevant, or be more responsive to the needs of:

- the academic curriculum;
- the teaching process; and
- the learning process.

The innovative intervention which will be used to illustrate the application of the model is an academic literacy course, Learning in English for Academic Purposes (LEAP), which was introduced into the Technikon curriculum for the first time in 1995. This skills-based course uses interactive teaching methodologies and co-operative learning strategies to develop academic literacy skills in first year students. The elements of the LEAP innovation which make it innovative in the Technikon context are the fact that: academic literacy courses do not form part of the mainstream Technikon curriculum; interactive teaching methodologies are seldom used as transmissive modes of delivery predominate; and co-operative learning strategies are uncommon in this context where learning is predominantly individual and often competitive.

While generalized models of evaluation are commonly found in the evaluation literature (e.g. Guba and Lincoln, 1989: 186–7; Madaus et al., 1983), it is this very general nature and often tacit acceptance of the context as an unchangeable given which limit their applicability for use within the constantly changing context of a South African education system in transformation. They do, however, provide a useful point of departure. Some fields, such as applied linguistics, have yielded detailed evaluation models (Rea-Dickins and Germaine, 1992; Mackay, 1994; Lynch, 1990) but they tend to be limited by the dictates of that particular field

and cover specifically programme/course evaluation. There is a dearth of evaluation models applicable to the transforming higher education context in South Africa, within which innovation takes place. The evaluation model presented in this article hopes to address this dearth. It emphasizes the need to locate the educational innovation which is being evaluated within the *context* and *policy framework* of its operation. The academic context surrounding educational innovation at any given educational institution and the policy framework governing the institution have a significant impact on innovative practices. Stage 1 of the model, which takes these factors into account, is thus seen as crucially important. It is also a stage which the aforementioned models lack, either in their endeavour to be too widely generalizable or too context specific. While the model is responsive to the specific dictates and evaluation needs of educational innovation within a context of transformation, it is flexible enough to be adapted for the evaluation of any kind of innovative educational practice.

Although the 10 stages making up this evaluation model are presented sequentially, the goals of the evaluation will determine whether the sequence of these stages is rigidly adhered to or used in a more cyclical fashion, as represented in the diagram in Figure 1 below.

Stage 1

This stage locates the innovation being evaluated within the context and policy framework of its operation. This stage is necessary at the outset of the evaluation process in order to fully understand the complex variables impacting on innovative practices. This is an especially important stage for the external evaluator for whom the context and policy framework are unfamiliar. Everitt (1995: 2), on evaluating public sector organizations and projects, claims that to evaluate practice without taking account of the **context** of that practice and the **policies** which constrain it or provide opportunities for it, assumes that practice exists as a commodity on its own that may be separated out for study (author's emphasis). Evaluating innovative educational practices as though they exist devoid of a context or accepting that the academic context and governing policies are unchallenged and unchangeable givens will serve to further exacerbate the potential that innovative practices have to impact on education systems. Decontextualizing evaluation, according to Everitt, serves also to remove policy, social structures and processes from critical scrutiny. Such scrutiny is imperative given the fact that the education system in South Africa is in a state of transformation. Unless the academic contexts within which educational innovation occurs and the policy frameworks which shape these contexts are seen to be part of the process of transformation and are challenged to change, there will be no significant reshaping of the existing education system in South Africa and innovative practices will continue to have minimal impact.

In the evaluation of the LEAP innovation at Peninsula Technikon, the overt goals of the innovation were outlined, then the innovation was contextualized by describing in detail:

- the **curriculum** being targeted (e.g. its intensity, structure, size, approach,

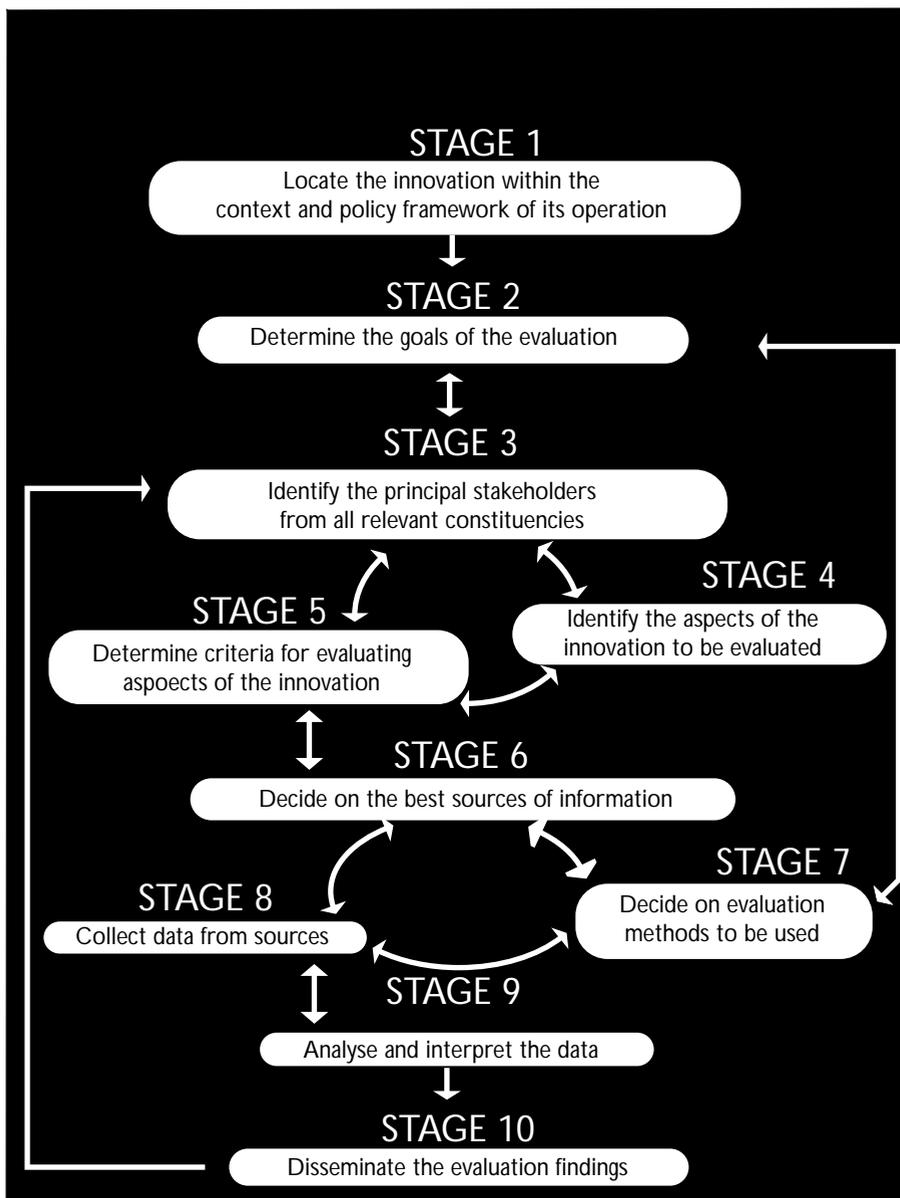


Figure 1 Stages in the Process of Evaluating Educational Innovation

content, learning processes, instructional materials, curriculum 'fit' and so on);

- the **teaching** being targeted (e.g. staff numbers, qualifications, familiarity with methodology, status of staff, contextual factors impacting on teaching such as funding and so on); and
- the **learning** being targeted (e.g. student selection process, language background, previous education, academic achievement, contextual factors impacting on learning such as time-tabling and so on).

Once the evaluation framework for the innovation has taken into consideration the specific levels of curriculum, teaching and learning, as well as all the broader levels surrounding these three processes, it needs to locate the innovation within the levels of the *academic context* of the particular educational institution and the state *policies* governing that education sector. An evaluator would therefore have to consider the effects of the broad institutional academic context and its governing educational policies on the innovation, and vice versa. Such evaluation considerations will inform a critical scrutiny of the constraints and opportunities operating at these two crucial levels of context and policy. They will also illustrate the potential of the innovation to shape and impact on these two levels.

In the evaluation of the LEAP innovation at Peninsula Technikon, the academic context and its governing policy framework were scrutinised and outlined by describing the opportunities and constraints provided by:

- the **institution**, e.g. its level of autonomy, policies (written) and practices (unwritten), prevailing institutional ethos/culture, financial/budgetary provisions and so on; and
- the **tertiary education sector**, e.g. state policies and practices, the prevailing social, economic and political climate in the country, and so on.

Evaluators of innovative practices, particularly at tertiary institutions in South Africa, have too often removed their objects of evaluation from the broad contexts and policy frameworks within which they operate. This not only removes the institutions and policies from the illuminative spotlight of evaluation but also raises them to a level beyond which they need to take responsibility for the challenges of transformation.

Stage 2

This stage determines the goals of the evaluation. This should not be confused with the goals of the innovation, which would have been described at the previous stage. The goals are usually determined by the source calling for the evaluation. Most evaluations are either internal, with formative goals and the purpose of improvement, or external, with summative goals and the purpose of accountability. The model suggests that evaluators of innovation should consider an eclectic approach by combining formative, summative and illuminative goals when planning an evaluation framework, irrespective of *where* the call for the evaluation has come from and *what* the evaluation needs of that particular stakeholder are. This eclectic approach is suggested in an attempt to satisfy as wide a

Evaluation 6(3)

range of stakeholders as possible and clearly link the goals of the evaluation to the needs of the various evaluation audiences.

The *formative* approach, commonly employed in the evaluation of innovative practice in South Africa, has as its purpose the improvement and increased effectiveness of the innovation. The *summative* approach has as its purpose accountability and judgement of impact, often using the demonstration of outcomes as evidence. Both of these approaches are commonly referred to in evaluation literature and used most often in evaluation studies. The *illuminative* approach, a term coined by Parlett and Hamilton (1972), is however seldom referred to and used less often in evaluation studies. This approach has as its purpose the illumination of innovations. Because innovation is often viewed with uncertainty, met with a resistance to change and often surrounded by misunderstandings in the broader academic context, there is a need for the illuminative approach which will shed light on the innovation and in this way clarify uncertainties, ease the broader academic community into the challenges of transformation and create a space for dialogue where misunderstandings can be voiced and shared understandings reached.

Evaluation studies seldom use these three goals together. This may be because a tension exists between the dual purposes of evaluation for improvement and accountability. The reflective nature of formative evaluation would highlight areas of weakness with the aim of improving them whereas summative evaluation would seek to highlight areas of strength in the hope of providing evidence of worth. The evaluator must be aware of this tension and guard against allowing one goal to undermine the other.

In the evaluation of the LEAP innovation at Peninsula Technikon, the formative evaluation goal served the participating teachers and learners who had the need for constant improvement and refinement of the innovation in accordance with their changing needs. The evaluation methods best suited to describing and understanding complex teaching and learning processes, probing possible areas of weakness and establishing the responsiveness of the innovation to the needs of this audience, were qualitative methods.

The summative evaluation goal of LEAP served the audience of policy makers to whom the innovators were accountable. This audience ranged from institutional policy makers to external funders and they had the need to make judgements about the worth of the innovation on which they would base policy decisions affecting future resourcing of the innovation. This audience generally undervalued qualitative methods as these were viewed as unscientific. They sought quantitative, empirical evidence of learning outcomes in terms of student achievement, untainted by the feelings and attitudes of those involved in the innovation.

The illuminative evaluation goal of LEAP served the entire academic community making up Peninsula Technikon. This audience needed to engage in discussions through which they could develop shared understandings of the LEAP innovation. The added bonus of these discussion forums was that they could be used by the innovators to promote the LEAP innovation and gain the support of the academic community. Because the purpose of the illuminative goal was to

open up debate around the attitudes and understandings of the academic community regarding the LEAP innovation, and because the purpose was for these people to share perspectives, perceptions and assumptions, qualitative evaluation methods were best suited to achieve that purpose.

Stage 3

This stage identifies the principal stakeholders from all the relevant constituencies at the institution. For a tertiary education institution the principal stakeholders would fall into the four broad categories of students, academic staff, institutional support services and policy makers. Once this stage in the evaluation process has been completed and all relevant stakeholders have been identified, the evaluator may wish to revisit Stage 2 of the model, as indicated in Figure 1. A clear understanding of the range of stakeholders might further clarify the goals of the evaluation since the audiences represented in Stage 2 will be constituted from among the stakeholders in Stage 3. The evaluator should thus move freely back and forth between Stages 2 and 3, allowing these two stages to inform each other.

Under the category of students, in the evaluation of LEAP a distinction was made between participants and non-participants in the LEAP innovation. The role and contribution of the LEAP participants was essential to the evaluation study as a whole, but more so in the formative phase. Although the non-participating students did not have direct exposure to the innovation, the possibility of their future participation in LEAP gave them a stake in the innovation too. The non-participating students also served as a useful pool from which to draw the control group required in the summative evaluation phase.

Under the category of academic staff a similar distinction was made between participants and non-participants in the LEAP innovation. The participating staff (such as teachers, trainers of teachers, curriculum developers) were the major contributors to the formative evaluation phase but the contribution of the non-participating staff was also valuable. Under this category, a distinction was made between lecturers who taught participating students but were themselves not directly involved in the innovation, and those lecturers who taught non-participating students. While all lecturers had a stake in the LEAP innovation, these two sets of non-participating lecturers had different interests in the innovation and consequently different contributions to make to the evaluation.

The category of support services contributed greatly to the illuminative phase of the LEAP evaluation. The support services at Peninsula Technikon were partially and in some cases wholly dependent on external funding, often undervalued by those who functioned in the mainstream, and viewed as convenient sites to bear the responsibility for institutional transformation. From this unfortunate position they had a clear stake in any innovation designed to meet the challenges of an institution in transformation.

The category of policy makers included those who were in decision making positions regarding the resourcing of the LEAP innovation, e.g. the rectorate, faculty heads, state departments, funding agencies. This group had a particular interest in the summative phase of the evaluation, since they constituted the

Evaluation 6(3)

powerhouse among the four broad categories and held the other three categories of stakeholders accountable.

Stage 4

This stage identifies the aspects of the innovation to be evaluated. These aspects should be determined collaboratively, involving input from as many stakeholders as possible as this process affords stakeholders a measure of control over the nature of the evaluation activity (Guba and Lincoln, 1989: 184). This cycle of collaboration, linking Stages 3, 4 and 5, is indicated in Figure 1. However, in keeping with the first stage of the model, two broad aspects are suggested for evaluation, namely the *innovation itself* (here the evaluator should consider the effects of the type of innovation and its impact on curriculum, teaching and learning) and the *institutional context* within which the innovation operates (here the evaluator should consider the social, political and economic climate impacting on the innovation). These aspects should, however, be negotiated through a cyclical process of collaboration.

Under the broad aspect of the LEAP innovation itself, the teaching was considered one of the more important aspects for evaluation. Here the teachers themselves, as well as the innovative methodology, were singled out for scrutiny. Under the broad aspect of the institutional context, the social climate was considered one of the important aspects for the evaluation of innovation. Here the range of institutional understandings of the LEAP innovation and the attitudes displayed towards it were singled out for scrutiny.

Stage 5

This stage determines the criteria for evaluating the aspects of the innovation which were identified at the previous stage. The criteria for evaluating educational innovations are determined very largely by the peculiarities of a specific innovation and the particular context within which the innovation operates. Criteria should be tailored to meet the needs of the specific evaluation study. There should also be a constant cyclical movement between Stages 4 and 5, as the determining of criteria will require the revisiting of aspects to be evaluated and vice versa. This cyclical movement also implies the continuing attempts by the evaluator to gain collaborative input regarding criteria for evaluating the various aspects of the innovation, from as many stakeholders as possible.

The LEAP teaching methodology (one of the aspects of the teaching process which was scrutinized) was evaluated in terms of *its effectiveness and extent of application*, as well as the *attitudes of participants* towards it. The institutional understandings regarding the LEAP innovation (one of the aspects of the institutional social climate which was scrutinized) was evaluated in terms of the understandings that a range of stakeholders had about *why, at which levels and how* LEAP should take place, as well as the *need that LEAP was serving*. The criteria by which all of the LEAP aspects were evaluated were determined collaboratively by the team of developers and informed by the threefold aims of the evaluation as well as the peculiarities of the innovation and its context. The criteria also evolved over the two-year span of the evaluation, as the aspects to be

evaluated were constantly being revisited and influenced by the preceding evaluation processes.

Stage 6

At this stage the evaluator decides on the best sources of information for evaluating the various aspects of the innovation using the criteria determined at Stage 5. Criteria may need to be re-assessed depending on the availability of sources, so there should be movement between Stages 5 and 6 as the one reshapes and informs the other. Just as the goals of the evaluation (determined at Stage 2) should serve the multi-faceted purposes of the evaluation, so too the sources of information should reflect the eclectic nature of an evaluation of innovative educational practice. As far as it is possible, given the timeframe and budgetary constraints of the evaluation, an attempt should be made to include all principal stakeholders (as identified at Stage 3) as sources of information in the evaluation.

In addition to the principal stakeholders, the evaluator becomes an important source of information, especially in the illuminative phase of the evaluation. Parlett and Hamilton (1972: 66) emphasize the role of the evaluator and the central place that observation occupies in illuminative evaluation. They stress the importance of building up 'a continuous record of ongoing events, transactions and informal remarks' through observing and documenting, as field notes, the day-to-day activities and environment surrounding the object of evaluation.

Besides utilizing stakeholders and personal observation records as sources of information, the evaluator should also consider non-stakeholder opinion regarding the innovation. Here the evaluator might approach educational experts and professionals outside of the institution where the innovation is occurring. Non-stakeholder opinion is valuable when cross-validating information from stakeholders in the interests of objectivity.

Another important source of information is documentation relating to the various aspects of the innovation being evaluated. Here the evaluator should analyse records ranging from institutional and state policy documents, statistics, survey profiles, funding proposals, correspondence, minutes of meetings and reports to instructional materials, student scripts and assignments, assessment criteria, marksheets, attendance records and other such course documentation.

Finally, a source seldom mentioned in the literature surrounding educational evaluation, but one that should be considered by the evaluator is the body of professional literature. Guba and Lincoln (1989: 211) cite this as a source from which additional information can be drawn but they caution that the information arising from a review of the literature deserves 'to be treated no differently from information gleaned from local documents and records, or from local observations'. The evaluator should therefore not elevate the knowledge contained in any body of literature to that of ultimate truths, but rather view the knowledge in much the same light as the information arising from the evaluation itself.

The way in which decisions are made about which sources to approach for what information will depend on the nature of the information which the evaluator is seeking. However, since most information reflects only people's perception of reality, the evaluator should attempt to gain information about any aspect of the

Evaluation 6(3)

innovation from more than just one source. This will ensure a more valid interpretation of the data at Stage 9. For example, in the LEAP evaluation, since the students were the only participants evaluating the teachers, the evaluator deemed it necessary to cross-validate this data with classroom visits for purposes of personal observation.

In the evaluation of the LEAP innovation, stakeholders supplied the bulk of the evaluative data although much of the data also came from documentation. All the 249 participating students were approached as sources, as well as an equivalent number of non-participating students who acted as a control group. All of the 17 participating academic staff members were approached as sources, as well as eight non-participating staff members. A sample of nine policy makers representing the different academic sectors at the institution was consulted, in addition to a survey of a wide range of institutional and state policy documents. Since the LEAP evaluation was an internal one, the evaluator, through her close involvement with the innovation itself and her familiarity with the context surrounding the innovation, became an important source of evaluative data as well.

Stage 7

At this stage the evaluator decides on the evaluation methods to be used. As with Stage 6, an eclectic approach is recommended but choices will ultimately be determined by a range of variables such as the type of innovation being evaluated, the timeframe and resources allocated to the evaluation, the sources of information available to the evaluator and so on. The type of information required by the different goals of the evaluation will also influence decisions about evaluation methods. A revisiting of Stage 2 is suggested here to enable the evaluator to reflect on the formative, summative and illuminative goals of the evaluation, as well as the qualitative and quantitative methods which were broadly linked to these goals at this early stage of the evaluation.

This stage also initiates the second cycle in Figure 1 which links Stages 6, 7 and 8. The evaluator will need to move freely between these three stages. For example, in an attempt to ensure congruency between data sources and methods of collection, the evaluator may see the need to change to a different methodology to suit the source; or, on analysing and interpreting data, a shortage of information may become clear and there may be a need to go back to the data collection stage.

In the LEAP evaluation, questionnaires calling for written answers were used where the sources were primarily second and third language speakers of English. An initial analysis of the questionnaire-generated data revealed the need to supplement this information with data generated through a different method. Oral interviews were then held with a representative sample of students to supplement the questionnaire data, as the interview method appeared to be more suited to the sources. The cycle of movement between Stages 6, 7 and 8 will continue until the final data interpretation has taken place. This final interpretation comes about after the evaluator has validated the findings with the sources of information and attempted to negotiate the interpretation with as many stakeholders as possible.

The importance of this process is emphasized by Stake (1985: 282) when he states that 'Negotiating drafts with key actors is more than a courtesy; it becomes essential to accuracy and completeness'.

The evaluator is advised to employ the best available method for meeting the evaluation criteria of each aspect of the evaluation. According to Parlett and Hamilton (1972: 64) 'the problem defines the methods used, not vice versa'. They further state that no method should be used exclusively or in isolation but rather that different methods be combined to throw light on a common problem. This use of triangulation techniques allows the evaluator to view the object of evaluation from a number of angles which in turn facilitates the cross-validation of data. Cohen and Manion (1980: 269–70) endorse this view by stating that 'exclusive reliance on one method may bias or distort the researcher's picture of the particular slice of reality he is investigating' and the more the methods contrast with each other, the greater the researcher's confidence in the findings. The use of contrasting methods, such as qualitative and quantitative, will also assist the evaluator in meeting the needs of the various audiences that the evaluation of an innovation should serve.

In the LEAP evaluation, predominantly qualitative methods were used for the formative and illuminative evaluation phases. For example, approximately 40 interviews were conducted, 500 questionnaires were distributed, personal observation of the teaching and learning processes was conducted, meetings and discussion forums were arranged and observed, documentation was scrutinized. Quantitative methods were employed for the summative phase and the evaluator made use of a pre- and post-test model with an experimental and control group. Statistics representing learning outcomes were also used.

Finne et al. (1995: 16), who present an action research model for evaluation including both formative and summative activities, argue that 'securing validity in evaluations is hardly enough, if the purpose is to make the evaluations useful'. They suggest that the evaluator goes beyond validity and considers what makes stakeholders trust evaluation results. This process, of ensuring the usefulness of the evaluation, they term 'credibility'. In their opinion, attention to this challenge will make the evaluation more credible to both the participants in the innovation being evaluated, as well as the outside stakeholders. In creating a credible evaluation the evaluator should ensure that each evaluation method used is accompanied by arguments for its validity, that attention is directed to an in-depth knowledge of the innovation which will enable the evaluator to make informed suggestions for improvement and that objectivity (so elusive in educational action research methodology) is substituted 'with honesty, critical distance, integrity and avoidance of conflict of interests'.

Macdonald (1974: 45), who favours the democratic evaluation model, highlights the importance of using data-gathering techniques which are accessible to non-specialist audiences, offering confidentiality to informants and giving them control over the use of the information they provide. This approach would have far-reaching implications for the evaluation model proposed in this article, where the evaluation serves such a wide range of audiences and where a multi-method approach to data-collection is advocated.

Stage 8

At this stage the evaluator collects data from the sources decided on at Stage 6 using the methods decided on at Stage 7. The interdependence of these stages at this fieldwork phase in the evaluation process is clear. Data collection instruments have to be designed taking into account the aspects to be evaluated, as well as the evaluation methods decided on and the particular sources of information. Measures should be taken to ensure the reliability and validity of all data collection instruments. For example, to increase the reliability and validity of a questionnaire designed for students who do not speak English as a first language, Pennington and Young (1989: 630) suggest that evaluation specialists are sensitive to the nature of the ESL context by providing 'opportunities for responses other than choices on rating scales' and orientating students 'to the content and purposes of the evaluation instruments and procedures'.

In the LEAP evaluation, the comprehensive questionnaire which was administered to the participating students included both responses to rating scales and written explanations for each of the questions. Students were also briefed as to the content and purpose of the questionnaire and facilitated through the completion of the instrument. This was necessary since the audience was generally unfamiliar with the process of course evaluation and questionnaire completion, and because some students required assistance in the interpretation of questions due to their poor levels of English proficiency.

The evaluator could also engage in pre-evaluation data collection where questions or instruments are tested with a sample of stakeholders to ensure their appropriateness. If time constraints do not allow for such pre-evaluation, the evaluator might need to refine or change evaluation questions as the data collection process proceeds. The LEAP evaluation, which covered a two-year period, refined evaluation questions as the evaluation was proceeding. Questionnaires were administered to all students mid-year and at the end of each of the two years. The questions appearing on the questionnaires were refined and modified after the first year on the basis of the data collected from students the previous year.

To distinguish each phase of data collection, Riley (1990) suggests that the evaluator should record the occasion, setting and method of collection. This was essential in the LEAP evaluation, especially during the illuminative phase, where unstructured recorded interviews resulted in about 20 audio tapes of data. After each phase in the data collection process the evaluator should reflect on the available information and allow each phase to inform and shape the next.

Stage 9

At this stage the evaluator analyses and interprets the data collected at the previous stage. The methods of analysis will be largely determined by the methods of evaluation determined at Stage 7. Other influencing factors, however, would be the number of respondents to a given instrument, as well as the evaluation need(s) being served by the data.

Qualitative data would lend themselves to methods such as discourse analysis, textual analysis and impressionistic interpretation: for example, in analysing an interview, the evaluator should look beyond the content to issues such as whether

all questions are answered, where the silent periods in the interview occur, which words frequently crop up and the non-verbal cues from the interviewee. If the qualitative data are extensive, the evaluator may need to employ the method of content analysis which converts the qualitative into quantitative, but, as Stake (1985: 281) points out, 'the uniqueness and contextuality of case data may be quickly lost' when this is done. The evaluation need being served by the data should then be considered. If the need is to 'increase subjective and intersubjective understanding among all stakeholders' (Vanderplaat, 1995: 89), a key concern of illuminative evaluation, then content analysis would not serve that need.

Quantitative data would lend themselves to statistical methods of analysis, although the limitations of this method for the evaluation of educational reform, innovation and transformation should be explained by the evaluator. Where possible the statistical analysis should be further illuminated with supporting qualitative data.

In the summative phase of the LEAP evaluation, where largely quantitative methods were used, qualitative methods were also employed as a means of cross-validating the summative data. This cross-validation of summative data is considered important by House (1980: 82-3) who claims that:

Quantitative argument should always be used in conjunction with human judgement, and human judgement should be given the superior position. Quantitative methodology should be seen to be based on human judgements and on intuitive reasoning and should be justified accordingly.

The experimental design, usually the basis for generating quantitative data, has moral and ethical implications for educational research as the control group is being denied the opportunity to experience the educational reform or innovation being evaluated. Further problems related to the use of experimental designs are that, in an attempt to deliberately control and manipulate the conditions being evaluated, the researcher creates an artificial 'laboratory' setting quite unlike the untidy reality surrounding innovative practices. The isolating of particular variables for scrutiny also creates a limited understanding of the total effects of any educational encounter. Beretta (1986: 153) suggests that 'what is required is a judicious balance between internal and external validity, between reliability and usability, and between certainty and relevance'. He further states, however, that the word causality should be deleted from evaluation vocabulary and that 'if true experiment is to legitimise causal statements, then true experiment is beyond the evaluator's reach' (p. 151).

After the data have been analysed and an initial interpretation made, the evaluator should return to as many stakeholders as possible to negotiate the findings. This phase is especially important for the illuminative goal of the evaluation, which seeks to generate debates around the innovation and create shared understandings. Everitt (1995: 7) charges the evaluator with the responsibility 'for ensuring that different views of participants in the practice, project or programme are not only expressed but are also heard by each other'. She sees the evaluator as the facilitator of this process as well as being tasked with providing evidence

Evaluation 6(3)

that will help inform disagreements and bring about resolution. This task would be best executed at this stage in the evaluation process.

Stage 10

At this stage the evaluator disseminates the evaluation findings. If, as a result of time and budgetary constraints, the evaluator has been unable to consult all stakeholders regarding the interpretation of the data, then at this final stage the evaluator should ensure that all stakeholders are informed of the evaluation findings, hence the cycle connecting Stages 10 and 3 in Figure 1. The presentation format could differ depending on the range and needs of the various stakeholders. For example, one may consider a formal evaluation report too lengthy and daunting to present to students and more applicable for funding agencies. A public forum may suit the needs of the broader academic community where there is an opportunity to ask and answer questions, whereas a formal presentation might be a format better suited to the needs of the institutional policy makers. Besides the format used for dissemination of the evaluation findings, the evaluator should also consider the way in which the information is conveyed to the different audiences. The goal of illumination and enlightenment could be undermined by factors such as complicated language usage, evaluation specific discourse and incoherent structure. Finally the evaluation findings need to be disseminated timeously or else the utility value of the findings will diminish.

Application of the Model

The decision to evaluate the LEAP innovation was intrinsically motivated. It was a decision taken by those directly involved in the materials development, as well as the implementation and teaching of the course. The evaluation was initially motivated by formative goals and a desire on the part of the materials developers to evaluate the materials they had produced and thereby improve them. This initial evaluation of the materials grew into a wider formative evaluation including aspects of the curriculum, teaching, learning and the actual model of intervention. The aim was to improve and increase the effectiveness of the aforementioned aspects for the audience of teachers and learners.

A call for a summative evaluation of LEAP came from the management of the institution. They were unconvinced by the subjective and introspective nature of the formative evaluation and required an objective form of evaluation using independent measures of outcomes by which they could judge the worth and effectiveness of LEAP and then effect policy decisions around the future of the innovation.

It was only at the end of the first year of implementation and after the first phase of the formative evaluation that the need for illuminative evaluation became clear. The developing team, in an attempt to market LEAP and gain institutional support for the innovation, encountered a lack of shared understanding at the institution about what exactly LEAP was, as well as a range of interpretations as to *how* and *where* institutional transformation should take place regarding the issues of language and learning. Data from the initial formative evaluation

also suggested that factors relating to the institutional context had influenced the success of the innovation. In an attempt to:

- place the institutional context under the scrutiny of evaluation;
- engage the academic community of the institution in debates around issues of transformation;
- develop shared understandings about LEAP; and
- inform policy decisions around the issue of transforming language and learning at the institution;

a decision to embark on illuminative evaluation was made.

Evaluation Outcomes

The formative phase spanned a period of two years. It addressed the evaluation needs of the five course developers, eight teachers and 249 (official course registration figures) learners, all of whom were involved in the evaluation process. The formative evaluation created a space for reflection on the processes of curriculum development, teaching and learning, and provided the various stakeholders with evaluative information on which to base future improvements.

The aspects of the LEAP innovation which were evaluated in this formative phase were determined collaboratively by the team of developers. It was decided to evaluate:

- **the model** (type) of intervention which LEAP represented, i.e. a stand-alone, skills-based, academic literacy course, by examining its strengths and weaknesses and comparing it to other models of intervention at the Technikon;
- **the curriculum** which the LEAP innovation targeted, namely the recurriculated credit-bearing English course for first year students at the School of Education, by examining the LEAP curriculum development process and the LEAP curriculum objectives, content, activities and instructional materials;
- **the teaching** of the LEAP course, by examining the teaching methodology promoted by LEAP as well as the LEAP teachers; and
- **the learning** as a result of students' interacting with the LEAP course, by examining the learning outcomes, student attitudes to learning and actual learning processes experienced in the LEAP course.

Since an eclectic approach had been adopted for the evaluation of LEAP, with the intention of satisfying the evaluation needs of as wide a range of stakeholders as possible, these stakeholders then became the major sources of evaluative data. Prominent evaluation theorists such as House (1980), Stake (1985) and Guba and Lincoln (1989), although they do not subscribe to any particular evaluation approach, all agree that stakeholder-based evaluation is crucial to an evaluation study which subjects the social context to scrutiny. A range of sources was thus consulted in the formative evaluation phase of the LEAP evaluation. The stakeholders considered most relevant for the formative evaluation were the

Evaluation 6(3)

students, teachers, trainers of teachers, curriculum developers and course moderator. Besides these stakeholders, another source used in the formative evaluation was a range of documentation relating to various aspects of the LEAP innovation, such as moderator's reports, review reports, student scripts and assignments, student journals, various evaluation sheets, informal meeting notes, LEAP course material and attendance records for LEAP classes and workshops.

These sources reflect the efforts to gain information about each of the various aspects of the LEAP innovation from more than just one source. This was done in the interests of a more valid interpretation of the data. Riley (1990) advocates that qualitative data should be collected in as many different ways and from as many different sources as possible. This concept of triangulation (Cohen and Manion, 1980), as it is referred to in evaluation literature, addresses the thorny issue of the validation of qualitative data. In an effort to validate data and consciously introduce the concept of triangulation, each aspect was evaluated using data from no fewer than three of the eleven different sources consulted. The stakeholders considered most relevant for the formative evaluation were those participating directly in the LEAP innovation. Students formed the majority of all participants and they were thus the major data source in the formative evaluation.

Essentially four qualitative methods were employed at the formative phase. These were: the questionnaire method; the interview method; personal observation; and analysis of documentation. Qualitative methods were favoured over quantitative methods at this stage of the evaluation, as the former method yields greater insights into the understandings and interpretations of stakeholders involved in the innovation being evaluated. Carr and Kemmis (1986: 103) refer to the failure of positivist approaches in recognizing 'the importance of the interpretations and meanings that individuals employ to make their reality intelligible'. Quantitative methods limit the respondents to illuminating predetermined criteria for evaluation, for example questionnaires 'eliciting responses to predetermined questions' (Cohen and Manion, 1980: 41), whereas qualitative methods, which are 'an exploration of processes and interactions' (Threadgold, 1985: 258), allow respondents to reflect and interpret their reality which is essential for the exploration of an innovation such as LEAP. While some questionnaires used in the formative evaluation, such as those distributed to the largest group of stakeholders, the students, did in fact make use of predetermined questions with fixed rating scales for answers, these were always balanced by allowing respondents the space to expand with further commentary. The overall results of the formative evaluation were very positive towards the innovation. All data sources recommended the continued implementation of the LEAP course and valuable insights were gained as to how the course might be improved for future implementation.

The summative phase primarily addressed the evaluation needs of the institutional policy makers, who required objective and independent measures by which the innovation could be judged. Only then were they prepared to effect policy decisions around the future of the innovation.

The summative phase focused on the learning outcomes aspect of the

innovation and examined the attainment of these outcomes, as well as the relationship between the innovation and the general academic performance of the participants. The data sources consulted for this phase were scores and statistics relating to the general academic performance of participating students in their mainstream courses, as well as the scores from an independent measure of learning outcomes. The summative evaluation goals were met using largely quantitative methods. A comparison of statistics representing learning outcomes in the LEAP course against other first year, mainstream courses, was one of the methods used in the summative evaluation. The pre- and post-test model, employing both experimental and control groups, was another method used for the summative evaluation. In this case a sample of LEAP students formed the experimental group, while a sample of Technikon first year students who were not participating in the LEAP course formed the control group. Both groups wrote the same pre-test and the same post-test. The post-test was different but equivalent to the pre-test, both of which tested the level of English language proficiency. The control and experimental groups were matched, as far as possible, in terms of home language, rural/urban origin and pre-test scores. These three variables were singled out for matching since they impacted most directly on the level of English proficiency. Trends in English proficiency testing prior to the summative evaluation had shown that English first language speakers had a distinct advantage over Afrikaans first language speakers, who in turn had a distinct advantage over African first language speakers. Also, urban students generally had an advantage over their rural counterparts. The results of the summative evaluation appeared to be inconclusive, as both the experimental and control groups showed no significant improvement in their post-test scores at the end of the first year of study. The experimental group was statistically the same sample as before, while the control group showed a significant deterioration.

While researchers should be aware that quantitative measures distort our understandings of reality, neither these measures nor the stakeholders calling for them can be ignored. When all the quantitative data from the summative phase of the evaluation were pooled, a consistent pattern of benefit seemed to emerge. All the summative data, such as the LEAP test scores, the LEAP throughput figures, the LEAP pass rates and the independent measure of outcome, showed that the LEAP participants had benefited as a result of their participation. However, the outcome of the LEAP summative evaluation seems to suggest that quantitative measures should be used circumspectly when evaluating educational innovation. Attempts at rigour often create other confounding variables. Researchers should display their awareness of these inherent limitations at the start of the evaluation and reflect on the kinds of quantitative data that can be used.

The evaluation instruments should also be planned and carefully designed before the innovation is implemented so that the summative data generated are justifiably measuring the desired outcomes. The appropriate qualitative data should then be obtained to inform and complement the quantitative data.

In the summative evaluation of LEAP, no attempt was made to control or manipulate conditions. Rather, the innovation was evaluated within the existing

Evaluation 6(3)

conditions and the researcher attempted to show how the existing conditions impacted on the apparent success or failure of aspects of the innovation. While there is benefit in reflecting reality in this way, the weakness lies in using existing measures which may not justifiably measure the desired outcomes. Evaluators need to take cognizance of this tension before embarking on an evaluation study.

The illuminative phase addressed the evaluation needs of the broad academic community at the institution and provided an opportunity for debate and discussion on the LEAP innovation. The aspects of the innovation which were evaluated in this phase of the process were the understandings and attitudes of the academic community towards the innovation. The sources consulted were all stakeholders previously mentioned in the formative phase, as well as institutional management and a range of institutional and state policy documents regarding institutional transformation. Qualitative methods were used during this phase, with unstructured interviews, open discussions, personal observation and the analysis of documentation predominating. Discourse analysis was employed at this phase of the evaluation as this method was best suited to disclosing understandings and underlying attitudes. This method analyses beyond simply *what* is said, by looking also at *how* it is said and what is left unsaid, as well as the context within which it is said. The biggest challenge to the evaluator was finding ways of ensuring that the different views of the various data sources were not only expressed but also heard by each other (Everitt: 1995). Interviews became a vehicle for achieving this purpose as there was a cross-pollination of ideas from one interview into another. The evaluator was able to test one interviewee's views against those of others who were interviewed later and in this way make cross-references. For example, misunderstandings regarding the nature of second language acquisition could be explored and clarified during interviews and different views regarding the locus of responsibility for language development could be carried from one interview to another and be debated.

Conclusion

The model provided the conceptual framework for the entire evaluation process. It enabled the evaluator to approach the evaluation in a structured and coherent manner, and facilitated the writing up and dissemination of the evaluation findings and conclusions. Stage 1, locating LEAP within the context and policy framework of its operation, engaged the evaluator in a deeper understanding of the institutional context, as well as the policies and practices that shape it. It ensured a close scrutiny of the range of contextual factors both constraining and enabling the LEAP innovation. It also provided a space, early on in the evaluation process, for reflection on the macro issues impacting on the innovation. This reflection provided insights that informed the direction of the evaluation, and brought about a keener understanding of the micro issues, relating to the LEAP innovation itself.

The second stage of the model, determining the goals of the evaluation, brought into sharp focus the evaluation needs of the respective audiences and the best methods to satisfy these disparate needs. It was a stage that flowed logically

from a scrutiny of the context and one that prepared the ground for the third stage, where the principal stakeholders were identified. The process of identifying stakeholders ensured that the evaluator considered and reflected their varying interests in the LEAP innovation, in the kinds of evaluative questions underpinning the enquiry. The following three stages, not rigidly sequential but rather each informing the other, highlighted the need for LEAP aspects which were evaluated to be congruent with the criteria for evaluation and the best sources of this evaluative information.

The final cycle in the model, especially Stage 7, which required the evaluator to revisit the goals of the evaluation, was particularly crucial in the LEAP evaluation. The LEAP evaluation had been motivated by formative goals at the outset, and this revisiting of the goals at Stage 7 drew the evaluator's attention to the evaluation needs of the broader audience of policy makers and the best evaluation methods to meet their needs. Although it was difficult to design the kinds of instruments that would best generate the data needed to measure the outcomes at this stage in the evaluation process, it did result in a more balanced evaluation which could speak to the disparate needs of the various audiences.

While the evaluation model proposed in this article places an onerous and immense responsibility on the evaluator(s), in so doing it also places the institutional context and its governing policies under the scrutiny of evaluation. In this way an innovation is not evaluated and understood in isolation, but rather as an integral part of the context which so powerfully shapes its very chances of success or failure.

References

- Beretta, A. (1986) 'Toward a Methodology of ESL Program Evaluation', *TESOL Quarterly* 20: 144-55.
- Carr, W. and S. Kemmis (1986) *Becoming Critical: Education, Knowledge and Action Research*. Geelong, Vic: Deakin University Press.
- Cohen, L. and L. Manion (1980) *Research Methods in Education*, 3rd edn. London: Routledge.
- Everitt, A. (1995) 'Developing Critical Evaluation', unpublished paper, Social Welfare Research Unit, University of Northumbria at Newcastle.
- Finne, H., M. Levin and T. Nilssen (1995) 'Trailing Research', *Evaluation* 1(1): 11-31.
- Guba, E. and Y. Lincoln (1989) *Fourth Generation Evaluation*. London: Sage.
- House, E. R. (1980) *Evaluating with Validity*. Thousand Oaks, CA: Sage.
- Lynch, B. (1990) 'A Context Adaptive Model for Program Evaluation', *TESOL Quarterly* 24 (1): 23-43.
- Macdonald, B. (1974) 'Evaluation and the Control of Education', in B. Macdonald and R. Walker (eds) *Innovation, Evaluation Research and the Problem of Control (SAFARI)*. Norwich: Centre for Applied Research in Education.
- Mackay, R. (1994) 'Undertaking ESL/EFL Programme Review for Accountability and Improvement', *ELT Journal* 48(2): 142-9.
- Madaus, G. F., M. Scriven and D. Stufflebeam (1983) *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Boston, MA: Kluwer-Nijhoff.
- National Commission on Higher Education (1996) 'A Framework for Transformation', unpublished report to the Minister of Education in South Africa.

Evaluation 6(3)

- Parlett, M. and D. Hamilton (1972) 'Evaluation as Illumination: A New Approach to the Study of Innovatory Programmes', Occasional Paper No. 9, Centre for Research in the Educational Sciences, Edinburgh.
- Pennington, M. and A. Young (1989) 'Approaches to Faculty Evaluation for ESL', *TESOL Quarterly* 23(4): 619-46.
- Rea-Dickins, P. and K. Germaine (1992) *Evaluation*. Oxford: Oxford University Press.
- Riley, J. (1990) *Getting the Most from Your Data*. Bristol: Technical and Educational Services Ltd.
- Stake, R. (1985) 'Case Study', in J. Nisbet and S. Nisbet (eds) *World Yearbook of Education 1985: Research, Policy and Practice*. London: Kogan.
- Threadgold, M. W. (1985) 'Bridging the Gap between Teachers and Researchers', in R. G. Burgess (ed.) *Issues in Educational Research, Qualitative Methods*. London: Falmer Press.
- Vanderplaat, M. (1995) 'Beyond Technique: Issues in Evaluating for Empowerment', *Evaluation* 1(1): 81-96.

CECILIA JACOBS works as a language co-ordinator in the Engineering Faculty, Peninsula Technikon, South Africa. She has previously lectured in English Communication Skills at the University of the Western Cape. Her research interests have been in the fields of second language acquisition and applied linguistics, learning theory, tertiary didactics and educational evaluation. Her current research focuses on the integration of language into specialist disciplines with a view to providing epistemological access to students for whom the medium of instruction is a second language.

Please address correspondence to: Department of Civil Engineering, Peninsula Technikon, PO Box 1906, Bellville 7535, Western Cape, South Africa. [email: jacobsc@abcnet.pentech.ac.za]