The Importance of Participatory Evaluation in Achieving Quality and Accreditation

Dr. Abdullah Al Fraidan
English Language Department
King Faisal University, Hofuf, Saudi Arabia

Abstract

Almost all academic institutions around the world – and in Saudi Arabia in particular – are looking to accredit both themselves and their academic programmes, and in accordance with this, The Department of English Language (DEL) at King Faisal University (KFU) is seeking to accredit its programme in English Language. One step towards achieving this goal would be to survey and review key aspects, notions, and systems in programme evaluation, especially those employed by others locally, regionally, and internationally. Therefore, this paper will shed some light on different methods of evaluation, and stress the importance of participatory evaluation in the quest to assure quality and achieve accreditation.

Keywords: Programme; evaluation; participatory; quality; accreditation; quantitative; qualitative.

Overview

Programme evaluation is a very important component in the process of education. Murphy (1985:2) in Rea-Dickins (1994) points out that 'Evaluation should be an integral part of the working of the curriculum to ensure that what is done is worthwhile, necessary, and sufficient'. However, there is a frequent tendency for evaluation to be left until the end of the programme – that is if it is integrated at all.

Yet, it cannot be ignored that evaluation is a very complex process that involves a lot of aspects. According to Lynch (2003:1), programme evaluation is 'the systematic inquiry into instructed sequences for the purpose of making decisions or providing opportunity for reflection and action.' This definition underlines the three main purposes of evaluation suggested by Rea-Dickens and Germaine (1992), which are: making decisions about the accountability (i.e. the effectiveness and efficiency of the programme); taking actions to better the curriculum; and providing opportunities for teacher development. Thus, the aims of evaluation do not only 'respond to demands of accountability but also … support institutional, curriculum and staff development.' (Rea-Dickins 1994:73). I would agree with Allison's argument (1999:57) that these purposes are not mutually exclusive. In addition to this, evaluation can be formative (i.e. ongoing/during the programme) and summative (at the end of programme), and it can be carried out by *outsiders* and/or *insiders.*

This paper offers a brief discussion on the quantitative and qualitative debate before examining the notion of participatory evaluation (PE) –to avoid repetition, PE is refered to her as participatory evaluation not programme evaluation- in relation to these two approaches. It will focus mainly on the feasibility, usefulness, and purposes of PE, and look at arguments both for and against this process. The paper will then conclude with a discussion on the value of the two approaches and the usefulness (or the otherwise) of PE and its application at KFU.

The Quantitative Approach

This is the traditional, scientific approach to data collection and analysis that attempts to verify some preferred viewpoints (Mackay et al. 1995). Another names for this approach is (logical) positivist (Oakley 2000). Two advocators of this approach are Suchman (1967) and Shadish (1995). It deals with either a true experimental or quasi-experimental design (Guba & Lincoln 1989, cited in Lynch 1996:15). In a true experimental design, subjects are randomly divided into two groups; one of these comes under the treatment, while the other is considered the control group and receives another programme. Pre-test(s) and post-test(s) would normally be included for both groups. In a quasi-experimental design, on the other hand, all the subjects receive the same treatment without having a control group. Lynch (1996) also mentions the non-equivalent control group where there is no random allocation of students.

The data is collected by measures that lend themselves to being turned into numbers and statistics. Examples might include test scores and the number of males and females in a programme (Brown 1989). The strengths of this approach lie in its 'meaningful, credible, trustworthy, accurate, and confirmable findings' through using the scientific method (Mackay et al. 1995). The findings can be easily analysed and interpreted, as this approach deals mostly with fixed items with predetermined information rather than open-ended questions, and usually there are no subjective decisions involved (Richards 2001). Weir & Roberts (1994) argue that it is more applicable to large scale studies, as is exemplified by both the Colorado (see Scherer, G.

and Wertheimer, M. 1964 for more) and the Pennsylvania (See Clarke, J.D. 1969 for more) projects, the aims of which were to evaluate the best foreign language teaching (F.L.T.) methods.

However, the quantitative approach suffers from the problems of control on the experiment variables and the lack of objectivity in the choice of the measures used to collect the data (Alderson 1994, Lynch 1996). Furthermore, the difficulty in randomising the samples and designing appropriate tests cannot be ignored. Thus, some researchers began to consider a move to the qualitative approach like Guba & Lincoln (1981) and Stake (1980).

The Qualitative Approach

Other names for this approach are 'naturalist' and 'interpretivist' (Oakley 2000, Lynch 2003). This method of analysis is naturalistic and less rigorous than the quantitative approach (Richards 2001), and seeks not to support any preferred viewpoints. According to Guba and Lincoln (1989), phenomena are best studied and understood in their context. The qualitative approach is 'exploratory, descriptive and discovery oriented in purpose' (Weir & Roberts 1994:159). It also gives a very detailed description of the learning process, attitude, and behaviour, and explains why the programme has worked or failed (Weir & Roberts 1994). It uses measures such as observations, diaries, interviews, journals, logs, documents, reports, and case studies, which usually cannot be expressed numerically (Richards 2001). The illuminative model, which can also be expressed in numbers when appropriate, is one example of those that use qualitative data.

There are, however, some problems with this approach. First, it suits small numbers, and second, it has the potential to generate huge amounts of information. This is why Lynch (1996) recommends focusing on specific purposes in order to overcome the latter problem. Indeed, the large quantity of information produced can make data coding and interpretation most difficult and problematic, and can also produce results that cannot be generalised to other situations. Furthermore, the qualitative approach lacks objectivity and credibility due to its reliance on subjective decisions, in which case a triangulation approach – i.e. gathering data from several sources and/or using different techniques (Lynch1996) – would be the ideal solution. Although there might be contradictory findings gathered from both techniques, one may clear the ambiguity of the other. Indeed, Lynch benefited greatly from combining the two approaches in his 'Reading English for Science and Technology' (REST 1992) project, in which he used qualitative data to clear the contradiction brought about by the findings of the pre- and post-tests. Correspondingly, the quantitative data clarified the contradiction of the learners' views towards grammar instructions. Lynch also claims that by mixing both paradigms he could reach very concise conclusions (Lynch 2003). Moreover, Weir and Roberts (1994) note that quantitative measures can be the start of gathering the qualitative data, and they see them as being complementary to each other. Other researchers, however, argue against the reconciliation of both quantitative and qualitative methods, highlighting 'cost efficiency and lack of procedures' as reasons for not using both (Lynch 1996:16) (see appendix 1 for a comparison between the two methods).

Moving on from this debate, the question that remains is who is going to *conduct* and *participate* in these methods?

Who?

In previous years, it has been preferable for evaluation to be conducted by 'outsiders', otherwise termed as 'JIJOES' which is an abbreviation of Jet-in-jet-out experts: they jet in for evaluating the programme in a short time then jet out (Alderson and Scott 1992) or fly-in-fly-out (Weir & Roberts 1994). For example, Elley (1989, cited in Silvester 1997) stresses the need for impartiality in evaluation, thus hinting at a preference for outsiders; similarly, Davies (1992) 'shows a preference for external evaluation' (cited in Allison 1999). On the other hand, Rea-Dickens and Germaine (1992) emphasise the importance of internal evaluations. I would argue that, ideally, external evaluations should be linked to internal evaluations.

Thus, we have explored in brief the weaknesses and strengths of the quantitative and qualitative approaches. What follows is a discussion concerning the notion of participatory evaluation, along with an exploration into its usefulness and feasibility, and an illustration of its advantages, disadvantages, and examples.

The Notion of Participatory Evaluation

The notion of PE is relatively 'new to the field of programme evaluation' (Lewkowicz & Nunan 1999:681). Some researchers have emphasised its importance and desirability; for instance, (Freire, 1970) Waters (1987) and Sharp (1989) state the importance of involving learners and teachers in course evaluation without mentioning any involvement of outsiders. Hopkins (1989:28) also hinted at this when speaking of evaluation as school improvement, in which 'the role of evaluator and the user are linked'. Moreover, Nunan (1989) urges for collaborative approach in the Australian Adult Migrant Education Program (AMEP), and researchers such as Alderson & Scott (1992), Alderson (1994), Mackay (1994, 1995) and Weir & Roberts (1994) have emphasised the importance of the participation of 'insiders' or 'stakeholders', and 'outsiders' in the evaluation process. Alderson (1994:2) believes that evaluation can be participatory 'when the decision making and planning of an evaluation, as well as its execution and reporting, is shared amongst all those involved in the project being evaluated, as well as by outsiders commissioned for the same purpose.' From this definition we see how the stakeholders must have an 'active role and involvement in the process of evaluation' (Mackay 1995), rather than standing apart, watching (Alderson & Scott 1992), and it distinguishes between those who plan and those who execute (Lilley 2005). Alderson also stresses that stakeholders should gain benefit, training, and experience from their involvement.

Although participatory evaluation is qualitative in its nature, I would argue that it can combine both quantitative and qualitative data, depending on the purpose. In a sense it is also illuminative when insiders highlight necessary information for the outsiders, or when both shed light on weaknesses and strengths for the decision makers. Furthermore, PE and other illuminative approaches are considered to be more democratic than the classical approach where stakeholders can express their points of view. In addition, PE is more formative than it is summative (Alderson & Scott 1992), but I would argue that it can sometimes combine the two (Lynch 2003), as there can be 'no evaluation purely summative nor purely formative' (Alderson & Scott 1992:39).

Purpose

It is my argument that PE can correspond to the three main purposes of evaluation: accountability, improvement, and teacher development. These purposes, however, can overlap; indeed, all three can be included in one project (Lynch 2003). Yet, it is obvious that participatory evaluation corresponds to the betterment purpose more than the others. This can be justified through a careful examination of Alderson's definition (1994), in which he stresses the benefits and experience gained by insiders' involvment. Therefore, although there are different aims and purposes for evaluation within each project, I would argue that the ultimate goal would be for insiders to develop their skills and expertise.

Usefulness

Calder (1994) and Alderson (1994) claim that the job of the outsiders cannot be completed without the cooperation of the insiders. It is my opinion that, whilst both outsiders and insiders could do the work without the assistance of the other, the results produced would not be as informative or reliable as those that would be obtained through working together. Indeed, teamwork such as this would enable the outsiders to train the insiders, thus improving the skills of the latter. Outsiders also can bring new perspectives (Mackay 1995), and accordingly, Lynch (1992, cited in Alderson & Beretta 1992) recommended that, instead of having just an insider perspective, the EST project in the University of Guadalajara should also involve an outsider, which might have led to more reliability and new findings.

When it comes to issues of time, PE can ease a lot of difficulties. For example, baseline data can often only be obtained by insiders (Weir & Roberts 1994), yet outsiders will need to gather it for their studies. Indeed, Lynch (2003:16) sees it as a mistake not to involve stakeholders from the primary level.

Furthermore, in instances where there are time restrictions or the need for improvement is urgent, a collaborative work would be the solution. Alderson and Scott (1992) emphasise the importance of allowing time for writing reports. This is something that can be done effectively using PE, as the task can be shared between both parties. In some cases, however, the availability of outsiders is limited (Alderson 1994), and in these instances insiders can do the 'donkey work' (Alderson & Scott 1992) whilst the outsiders observe, legitimate the work, and help with reporting and analysing.

Moreover, when stakeholders are involved in the process their willingness towards the project will increase, which subsequently increases the readiness. This, according to Mackay (1995), is a key to success. Such involvement can also create an enthusiastic, collaborative environment as well as increase the stakeholders' awareness of what is going on in the programme. Furthermore, it may lead to self-evaluation, then to action research, and then to self-improvement. Equally important is the 'ownership' feeling – that is the degree to which the participants feel that the innovation 'belongs' to them (Kennedy 1988). Kennedy believes that ownership is an important condition in evaluation for change; however, ownership alone is not enough to ensure success (Lilley 2005).

In some circumstances, PE has led to a change in attitude. For instance, in the Geneva project (1994) the stakeholders' attitude changed from one of 'no involvement' to 'demanding full

control'. Similarly, in Morrow & Schocker's (1993) INSET project, teachers' attitudes were seen to change and they began to consider transferring their experience to their own contexts.

Furthermore, PE can satisfy both bureaucratic and democratic demands, as well as fulfilling administrative needs. The findings of a given situation can be used as the basis of another programme, and although PE is situation-specific, its findings might be germane to other similar contexts.

Feasibility

The apparent feature of PE is that it is democratic, and Waters (1987) argues that PE can only be feasible when there is democracy in the evaluation context. Thus, it cannot be applicable when the educational system is very bureaucratic and does not allow stakeholders to participate in decision making. Waters goes on to claim that PE (involving learners in ESP contexts) is a risky business, but is not impossible.

I agree with Alderson and Scott's argument (1992) that to constitute a PE model two conditions must be met: there should be an active involvement of insiders in almost all stages, and they must benefit from the process. Alderson and Scott also insist that the need for training should not weaken participation.

Furthermore, we must not forget that PE is likely to take a lot of time and effort. Moreover, sometimes it may not suit the initial stages of evaluation if the stakeholders have not been clearly identified from the beginning (Lewkowicz & Nunan 1999). Worse still, insiders can perceive outsiders as being a threat. Nonetheless, this misunderstanding can be avoided in different ways, such as ensuring the early involvement of insiders, and articulating the aims of the project to ensure familiarity between both parties. However, this familiarity can sometimes prevent the insiders from freely expressing their opinions (Silvester 1997). Furthermore, we must give consideration to the cultural mismatch between insiders and outsiders (Kennedy 1988), as this may either cause conflict or lead to change in the target culture.

Examples of some International Projects

In order to gain insight into the best procedures for designing a method of programme evaluation, it is important to look at the experiences of others who have formulated processes in the past. In doing so, we may benefit by adopting those methods proved to be strong, and overcoming those shown to be weak.

Alderson & Scott 1992: The ESP Project in Brazil

This project came at a time when there was an intrinsic desire for evaluation in the Brazilian Federal Universities, and it was a much-publicised attempt to involve stakeholders in the process.

Alderson and Scott emphasised the importance of involving as many participants as possible within the seven stages of the evaluation in order to ensure success and minimise any feeling of threat between the participants (see appendix 2 to learn who was involved in each stage). This project was considered to be a success, although there were some problems encountered, namely: the length of time it took; the stakeholders' lack of training; inadequate and incomplete data

collection; difficulty in data interpretation; inadequate sampling; and lack of classroom observation. Berretta (1992) also mentions the 'high price to pay' for such an evaluation. However, Alderson and Scott (1992) believe that PE is '*possible and rewarding*'.

Williams & Burden 1994: The Geneva Project

This project took place in a primary school in Geneva in which the headmaster wanted to run an 'immersion' social studies programme, and thus sought help. The outside evaluators chose an illuminative approach enabling questions and problems to arise naturally. Quantitative and qualitative data were used and discussed in two phases. The results were promising in the sense that the stakeholders' attitudes changed; where previously their misunderstanding of the project's aim had led them to perceive it as being a threat, they developed more of a desire to own the evaluation, and also recognised the need for cooperation.

Mackay et al. 1994: The Bali Project by ELTPU

ELTPU stands for English Language Teaching Projects Unit. In this project involving the participation of 25 schools in Indonesia, Mackay et al. (1994) used a collaborative model. The aims of the project were to assess the centres' practices and activities, and to ensure maximum quality of the programmes. A performance indicator framework was used, which was agreed on by all of the stakeholders without being imposed on them (see appendix 3). This framework consisted of *units* (e.g. institutes), which were then divided into *foci* (e.g. programmes). Each focus was broken down into *key areas* (e.g. quality of each course) which in turn were given *performance indicators* (e.g. for each course: complete documentation). These indicators were then broken down into 'themes'. Collected information was compared against these indicators in order to assess the level of weakness or strength in each theme. In a sense, this method resembles Stake's countenance model (1980, See Hopkins (1989) for more about this model).

The project managers established that for this collaborative model to work, it would require minimal training of its participants in order to ensure their involvement. The project included an investigation into what *could* be improved and which matters were urgent. All in all, the project was considered a success.

Lewkowicz & Nunan 1999: The Hong Kong IELP Project

IELP stands for Intensive English Language Programme. This project took place in the secondary-school system in Hong Kong with multiple stakeholders. It lasted for three years, focusing on different aspects in each year. The stakeholders were not involved from the beginning because they were not identified in advance. The aim of the project was to introduce a bridging programme to change the medium of instruction from Chinese into English. Similar to the ESP Brazil project, there was an intrinsic motivation for evaluation aroused from the teachers' classrooms. They used a model that was characterised by five principles (see appendix 4). Both quantitative and qualitative data were used, gained through questionnaires, classroom observation, and written language samples. These findings were detailed in a draft and then a final report. Although the project was considered a success, evaluators were faced with a number of problems. Some of the teachers were recruited and thus had no interest in the project, therefore levels of motivation differed. Personnel kept changing, and there was a general lack of appreciation of the model's five principles. Furthermore, some participants saw it as being the

evaluators' job to decide how to proceed in each stage, and stakeholders regarded the outsiders as a threat, and the workshop discussion as criticism.

Therefore, whilst Lewkowicz and Nunan argue that a participatory model was the most appropriate one for the project, a *true* collaboration was not possible due to the problems faced.

Advantages and Disadvantages of Participatory Evaluation

The inherent problem with PE – and one that has caused a lot of criticism – is the issue concerning the time it takes to carry out the work and collect the data, and this is especially apparent when comparing it to the JIJO model. In addition, the 'high price to pay' (Beretta 1992) and data analysis can cause difficulties and be time consuming.

Sometimes insiders rely on outsiders for some of the work as they are the experts (e.g. the Hong Kong project). Outsiders are considered more objective, clear-sighted, of a higher status, and more experienced – but not always necessary (Alderson & Scott 1992). However, Wolter (2000) notes that problems can be avoided by using the participants' knowledge of the local learning/ teaching situation. Yet, some teachers would consider such involvement to create an extra workload, especially if they are poorly paid (Kennedy 1988). On the other hand, insiders are more familiar with their context. They are more reliable, and are better at evaluating the worthiness of their programme than outsiders. Furthermore, they take action despite their lack of skills in data collection and making recommendations. PE offers them a chance to increase their skills, which might lead to self-evaluation and bring about improvements. Insiders also have no investment, thus they are less biased. Therefore, the question that remains is ... why are outsiders still more favoured? (Alderson & Scott 1992).

The KFU Institutional Accreditation Project (11/11)

This project was given the motto '11/11' after the date (11/11/2011) that was set for the external reviewers to make a decision about KFU's eligibility to achieve institutional accreditation.

KFU made a previous attempt at this in 2008 and external reviewers issued around 80 recommendations. In 2011, other external reviewers checked the university's practices and found that it fulfilled more than 75% of the previous recommendations. One reason for this was the implementation of a partial notion of participatory evaluation. KFU suffered from a low number of responses to students' surveys. I suggested introducing incentives for completing the surveys, such as ipads and iphones, and after this the rate of response started to increase. Staff members' surveys are not open to the public yet, and their voices have started to reach higher administrations.

The DEL at KFU decided to look for an accrediting agency to review its BA programme in English Language, which includes four domains (language skills, linguistics, translation, and Literature). To pass the review successfully, the DEL needs to incorporate all of the models and criteria set up by others (see appendices), as well as to demonstrate that it has benefited from its last two experiences.

Conclusion

Although the quantitative approach is scientific, saves time and produces more reliable findings, it does not describe the process of the evaluation, which is a major problem. Moreover, Mackay (1994:314) claims that evaluation is not 'an experiment nor a controlled group'. The qualitative approach, on the other hand, lacks credibility due to its large element of subjectivity. It also takes a lot of time and there are difficulties involved in coding the data. Mackay (1994) argues that subjectivity should be minimised in both approaches. However, we cannot single out one method as being superior, as the suitability of each would depend on its purpose (Weir & Roberts 1994). Therefore, giving preference to just one method over the other would be 'inappropriate and unrealistic' (Alderson 1994:13). Ideally, both techniques should be involved in any evaluation, as evaluators would benefit greatly from both, especially in a participatory model.

Of course, PE still has the potential problems of the 'high price to pay' and the large time frame that it requires. Nevertheless, Rea-Dickens and Germaine (1998) point out that PE eases data collection, contributes to the participants' commitment to the changes being evaluated (ownership), and enhances professionalism within the field. Therefore, the KFU's Department of English Language, and indeed all the other sectors within the university, should target a participatory evaluation in their accreditation project.

Ultimately, participatory evaluation is considered to be the most satisfactory model and one that is superior to others (Alderson 1994). As Alderson and Scott (1992) note, it fits the Chinese saying, 'Tell me, I forget; show me, I remember; involve me, I learn.'

About the author

Dr. Al Fraidan received his MA and PhD from Essex University, UK. He is working as an assistant professor in the English Language Department at King Faisal University, Saudi Arabia since 2010 up to now. He is holding three positions at King Faisal university: Vice-dean of E-learning & Distance Education Deanship, Head of the English Language Department, Director of the English Language Centre.

References:

Alderson, C. and Beretta, A. (1992) (eds.). *Evaluating Second Language Education*, Cambridge University Press

Alderson, C. and Scott, M. (1992). Insiders, Outsiders and Participatory Evaluation in C. J. Alderson and A. Beretta, (eds.) 25-60

Alderson, C. (1994). Towards a Theory of Participatory Evaluation: Insights from Applied Linguistics in L. Barbara and M. Scott (eds.). *Reflections on Language Learning*, Clevedon: Multilingual Matters, 1-13

Allison, D. (1999). *Language Testing and Evaluation: An Introductory Course*, Singapore University Press, 57-71

Bennett, J. (2003). *Evaluation Methods in Research*, Continuum

Brown, J. D. (1989). Language Program Evaluation: A Synthesis of Existing Possibilities in R. K. Johnson (ed.). *The Second Language Curriculum*, Cambridge University Press

Calder, J. (1994). *Programme Evaluation and Quality*, London: Kogan Page Ltd.

Celani, M. A., Holmes, J. L., Ramos, R. C. G. and Scott, M. R. (1988). *The Brazilian ESP Project: An Evaluation*, Sao Paulo: Editora de PUC-SP

Clarke, J.D. (1969). The Pennsylvania Project and the "Audio-lingual vs. Traditional" Question', *The Modern Language Journal*, *53* (6)

Davies, A. (1992). Review of C. J. Alderson & A. Beretta (eds.). *Language Testing*, *2* (2) 207-209

Elley, W. B. (1989). Tailoring the Evaluation to Fit the Context, in R. K. Johnson (ed.). *The Second Language Curriculum*, Cambridge University Press

Freire, P. (1970). *Pedagogy of the oppressed*. Continuum

Guba, E. G. and Lincoln, Y. S. (1981). *Effective Evaluation*, San Francisco: Jossey-Bass

Guba, E.G. and Lincoln, Y.S. (1989). *Fourth Generation Evaluation*, Newbury Park, CA: Sage

Hopkins, D. (1989). *Evaluation for School Development*, Open University Press, pp. 18-28

Kennedy, C. (1988). Evaluation of the Management of Change in ELT Projects, *Applied Linguistics*, *9* (4) 229-342

Lewkowicz, J. A. and Nunan, D. (1999). The Limits of Collaborative Evaluation, *TESOL Quarterly*, *33* (4). 681-700

Lilley, T. (2005). *Language Programme Evaluation Course Handouts*, University of Essex, weeks 17-25

Lynch, B. K. (1992). Evaluating a Program Inside and Out, in C. J. Alderson and A. Beretta (eds.) 61-99

Lynch, B. K. (1996). *Language Program Evaluation: Theory and Practice*, Cambridge University Press

Lynch, B. K. (2003). *Language Assessment and Programme Evaluation*, Edinburgh University Press

Mackay, R. (1994). Undertaking ESL/EFL Programme Review for Accountability and Improvement, *ELT Journal*, *48* (2), 142-149

Mackay, R., Wellesley, S. and Bazergan, E. (1995). Participatory Evaluation, *ELT Journal*, *49* (4) 309-317

Morrow, K. and Schocker, M. (1993). Process Evaluation in an INSET Course, *ELT Journal*, *47* (1) 47-55

Murphy, D. F. (1985). Evaluation in Language Teaching, in C. J. Alderson (ed.). *Evaluation*, Lancaster Practical Papers in English Language Education*, 6,* Oxford: Pergamon Press .1-17

Nunan, D. (1989). Towards a Collaborative Approach to Curriculum Development: A Case Study, *TESOL Quarterly*, *23 (*1), 9-25

Oakley, A. (2000). *Experiments in Knowing*, Cambridge: Polity Press

Rea-Dickens, P. and Germaine, K. P. (1992). *Evaluation*, Oxford: Oxford University Press

Rea-Dickens, P. (1994). Evaluation and English Language Teaching, *Language Teaching*, *27*, 71-91

Rea-Dickens, P. and Germaine, K. P. (eds.) (1998) *Managing Evaluation and Innovation in Language Teaching,* Longman

Richards, J. (2001). *Curriculum Development in Language Teaching*, Cambridge University Press, 286- 309

Scherer, G. and Wertheimer, M. (1964). *A Psycholinguistic Experiment in Foreign Language Teaching*, New York: McGraw-Hill

Shadish, W. R. (1995). Philosophy of Science and the Quantitative-Qualitative Debate, *Evaluation and Program Planning*, *18* 63-57

Sharp, A. (1989). Staff/ Student Participation in Course Evaluation: A Procedure for Improving Course Design, *ELT Journal*, *44*(2) 132-137

Silvester, H. (1997). INSET, Projects and Evaluation, in D. Hayes (ed.). *In-service Teacher Development: International Perspectives*, Prentice Hall, 109-115

Stake, R. E. (1980). Program Evaluation, Particularly Responsive Evaluation, in W. B. Dockrell and D. Hamilton (eds.). *Rethinking Educational Research*, London: Hodder & Stoughton. 72-87

Suchman, E. (1967). *Evaluation Research*, New York: Russell Sage Foundation.

Waters, A. (1987). Participatory Course Evaluation in ESP, *ESP Journal*, *6* (1) 3-12

Weir & Roberts (1994). *Evaluation in ELT*, Oxford: Basil Blackwell

Williams, M. and Burden, R. (1994). The Role of Evaluation in ELT Project Design, *ELT Journal, 48*(1) 22-27

Wolter, B. (2000). A Participant-Centered Approach to INSET Course Design, *ELT Journal*, *54(*4), pp. 331-318

Appendices:
Appendix 1: A table showing the differences between the quantitative and qualitative approaches (Oakley 2000 cited in Bennett 2003).

Oakley's summary of the two main paradigms of educational research and evaluation

	'(logical) positivist'/'scientific'/'quantitative'/ 'positivism'	'naturalist'/'interpretivist'/'qualitative'
Aims	Testing hypotheses/generalizing	Generating hypotheses/describing
Purpose	Verification	Discovery
Approach	Top-down	Bottom-up
Preferred technique	Quantitative	Qualitative
Research strategy	Structured	Unstructured
Stance	Reductionist/inferential/hypothetico-deductive/outcome-oriented/exclusively rational/oriented to prediction and control	Expansionist/exploratory/inductive/process-oriented/rational and intuitive/oriented to understanding
Method	Counting/obtrusive and controlled measurement (surveys, experiments, case control studies, statistical records, structured observations, content analysis)	Observing (participant observation, in-depth interviewing, action research, case studies, life history methods, focus groups)
Implementation of method	Decided *a priori*	Decided in field setting
Values	Value-free	Value-bound
Instrument	Physical device/pencil and paper	The researcher
Researcher's stance	Outsider	Insider
Relationship of researcher and subject	Distant/independent	Close/interactive and inseparable
Setting	'Laboratory'	'Nature'
Data	Hard, reliable, replicable	Rich, deep, valid
Data type	Report of attitudes and actions	Feeling, behaviour, thoughts, actions as experienced or witnessed
Data analysis	Specified in advance	Worked out during the study
Analytic units	Predefined variables	Patterns and natural events
Quality criterion	Rigour/proof/evidence/statistical significance	Relevance/plausibility/illustrativeness/responsiveness to subjects' experiences
Source of theory	*A priori*	Grounded
Relationship between theory and research	Confirmation	Emergent
Causal links	Real causes exist	Causes and effects cannot be distinguished
Nature of truth statements	Time- and context-free generalizations are possible	Only time- and context-bound working hypotheses are possible
Image of reality	Singular/tangible/fragmentable/static/external	Multiple/holistic/dynamic/socially-constructed
Research product	Stresses validity of research findings for scholarly community	Stresses meaningfulness of research findings to scholarly and user communities

Taken from Oakley, 2000, 26–7.

Appendix 2: A table of the seven stages in the Brazil Project, showing who was involved in each stage (Alderson & Beretta 1992:42)

PARTICIPATION SUMMARISED

	insiders				outsiders	
	UT	UC	PC	RA	EC	SP
designing the whole evaluation	﹡	﹡	﹡﹡	﹡	﹡﹡	
constructing the instruments	﹡﹡	﹡﹡	﹡﹡	﹡﹡	﹡﹡	
testing them out	﹡﹡	﹡﹡	﹡﹡			
collecting data	﹡﹡	﹡﹡	﹡﹡	﹡﹡		
data analysis	﹡﹡	﹡﹡	﹡﹡	﹡﹡	﹡﹡	
drafting the report		﹡﹡	﹡﹡	﹡﹡	﹡	
reading and learning from it	?	?	﹡﹡	?	﹡﹡	﹡﹡

Key: ﹡ = approximate degree of involvement
? = no evidence available
UT – University Teachers
UC – University Co-ordinators (who are also UTs)
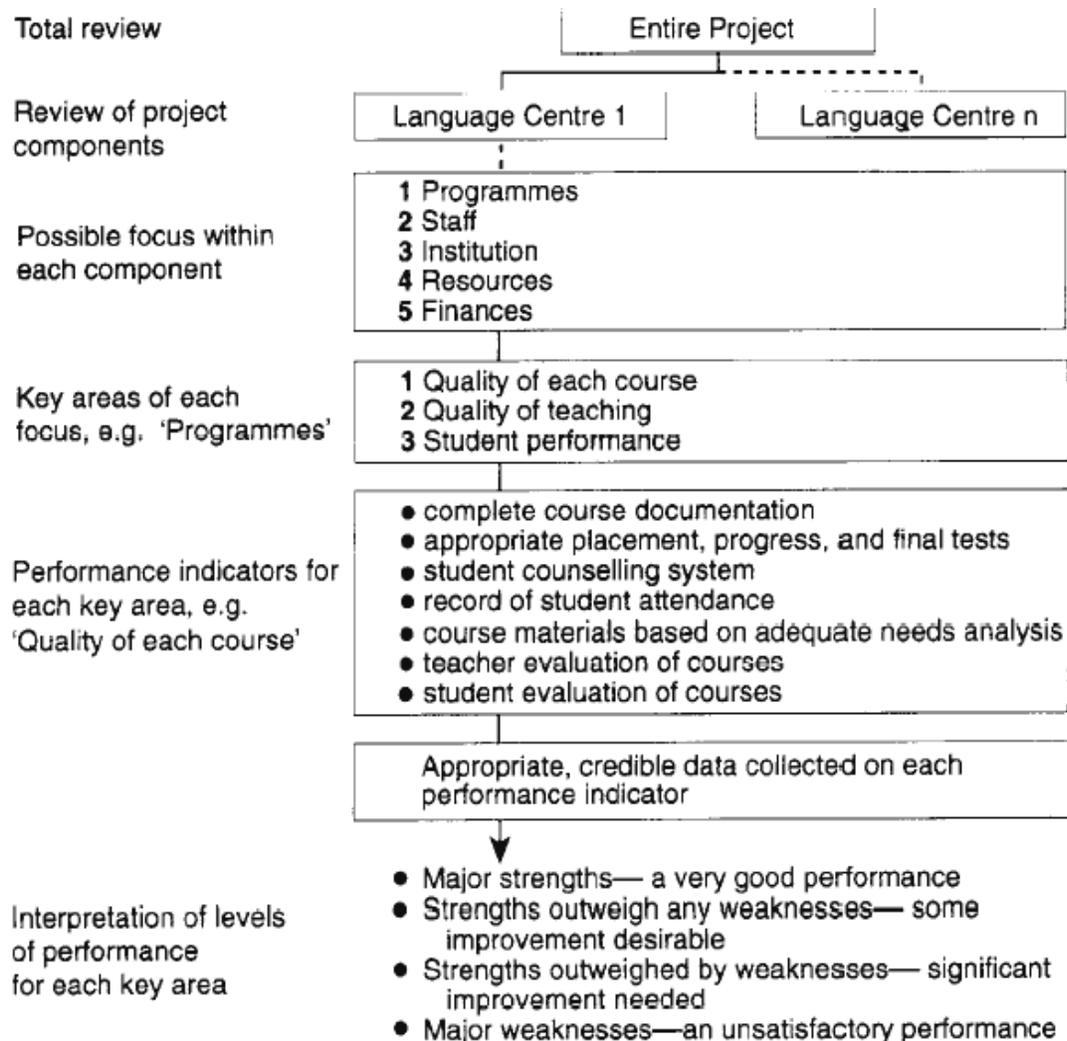PC – São Paulo Project Co-ordinators
RA – Research Assistants in CEPRIL (stay approx. 2 years)
EC – External Consultant
SP – Sponsors
A fuller description of these stages is to be found in Celani *et al.* (1988) especially Chapter 3.

Appendix 3: The performance indicator framework used in the Bali Project (Mackay 1994:147)

Total review	Entire Project
Review of project components	Language Centre 1 Language Centre n
Possible focus within each component	1 Programmes 2 Staff 3 Institution 4 Resources 5 Finances
Key areas of each focus, e.g. 'Programmes'	1 Quality of each course 2 Quality of teaching 3 Student performance
Performance indicators for each key area, e.g. 'Quality of each course'	• complete course documentation • appropriate placement, progress, and final tests • student counselling system • record of student attendance • course materials based on adequate needs analysis • teacher evaluation of courses • student evaluation of courses
	Appropriate, credible data collected on each performance indicator
Interpretation of levels of performance for each key area	• Major strengths— a very good performance • Strengths outweigh any weaknesses— some improvement desirable • Strengths outweighed by weaknesses— significant improvement needed • Major weaknesses—an unsatisfactory performance

Appendix 4: The Five Principles Model used in the Hong Kong Project (Lewkowicz & Nunan 1999:688)

1. Negotiate a set of procedural principles for the conduct of the evaluation. In line with Rea-Dickins and Germaine (1998), we believed that this stage was most important because, as evaluators, we were keen on enhancing the commitment of all those involved in the evaluation process to the principles of the evaluation. We also believed that such negotiation would help reduce the "them-and-us" dichotomy that often arises when evaluators are brought in from outside (the JIJOE syndrome).

2. Develop an evaluation brief in collaboration with the stakeholders. Because the evaluation was to last for 3 years, it was important to determine how and when the data were to be collected and to elicit the help of the course directors and teachers in collecting the data; again, this step was necessary if stakeholders were to be committed to an evaluation, according to Rea-Dickins and Germaine (1998). At the outset we recognized that data were to be collected selectively over the 3 years and that everyone concerned needed to understand the focus of each phase of data collection.

3. Collect data from all stakeholders. To fully appreciate what was going on during the teaching and to understand why some students who had the opportunity to enroll failed to do so, we needed to obtain data from all stakeholders. We also needed to view all groups of stakeholders as equally important in providing input for the data-gathering process.

4. Negotiate the content of reports and recommendation. Again, we believed that the evaluation should be seen not as top-down but as one involving a number of parties.

5. Provide feedback to the participants in the form of interactive workshops. Such workshops, though possibly novel in their approach, would in our opinion provide an opportunity for dialogue and for planning future stages of the evaluation, which was important for a long-term project.