Language Teaching Research

http://ltr.sagepub.com/

Small answers to the big question: Learning from language programme evaluation

Richard Kiely

The online version of this article can be found at:

http://ltr.sagepub.com/content/13/1/99

Published by:

$SAGE

http://www.sagepublications.com

Additional services and information for *Language Teaching Research* can be found at:

Email Alerts: http://ltr.sagepub.com/cgi/alerts

Subscriptions: http://ltr.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://ltr.sagepub.com/content/13/1/99.refs.html

Small answers to the big question: Learning from language programme evaluation

Richard Kiely *The University of Bristol, UK*

This paper explores why the learning posited as an intrinsic dimension of evaluation practice and use has been difficult to achieve, and how it might be more effectively realized. In recent decades language programme evaluation has evolved from focused studies of teaching methods inspired by language learning theories to a curriculum management enterprise with a focus on quality assurance and enhancement. There has been a parallel development from evaluation as a research-type study, where the findings are disseminated through publications, to evaluation as a dialogue within programmes for ongoing improvement of learning opportunities. These changes raise issues about learning from evaluation. Who should learn from programme evaluation? What should they learn? And how? The argument in this paper is for a synthesis of these conceptions of programme evaluation, so that the research-type knowledge-building enterprise and the ongoing quality management processes are mutually informing, and programme evaluation becomes a socially-situated cycle of enquiry, dialogue, and action. To facilitate the integration of research and management perspectives, three aspects of programme development are explored: innovation, teachers at work, and the quality of the learning experience of students. These are aspects of programmes which have not always had attention in evaluation theory and practice, but are central to understanding how language programmes work and develop. The significance of these issues is explored in episodes from the evaluation of materials used in an English for Academic Purposes (EAP) programme.

Programme evaluation is a form of enquiry which describes the achievements of a given programme, provides explanations for these, and sets out ways in which further development might be realized. The big question in the title of this paper is about programme effectiveness, and traditionally the answer has been sought in terms of test results, a language learning theory, or a particular syllabus. However, trustworthy answers have not been forthcoming, and now the challenge is to document the range of actions, reactions, and contributions that shape a programme. The task is thus a broad, holistic one, incorporating all aspects of the programme and informed by all stakeholders. The embedding of evaluation in professional practice within programmes and realizing of the central role of stakeholders are directions posited by Rea-Dickins (1994):

Richard Kiely, Graduate School of Education, University of Bristol, BS8 1JA, UK; email: R.Kiely@bristol.ac.uk

> If evaluation in English Language Teaching is to be effective, we will see a stronger inte-
> gration of evaluation within practice, as part of an individual's professionalism and an
> increase in collaborative activity where teachers (and other relevant participants) are
> actively engaged in the monitoring process. (p. 84)

Pennington (1998) adds a process dimension, *interaction*, to this notion of
evaluation as a dimension of professional action. Norris's (2006) recent char-
acterization also extends the conception of programme evaluation by iden-
tifying four distinct purposes, reflecting the traditional task of evaluation,
judging, the current focus on quality assurance, *demonstrating*, the know-
ledge-building enterprise, *understanding*, and the challenge of programme
development, *improving*. The central theme of these rippling, inclusive views
of evaluation is the realization of the potential for learning; that is, stake-
holder understanding of how language programmes work in particular con-
texts and can be improved. Stakeholders include programme participants
(students and teachers), policy makers and sponsors (programme and institu-
tional managers, accrediting bodies), and those who may benefit from evalu-
ation processes, findings, and outcomes (language learning and teaching
researchers, teacher educators, and the evaluation community) (Kiely & Rea-
Dickins, 2005).

There are three issues however, which may constrain this learning. First,
there are likely to be tensions between the different purposes. For example, the
task of *improving* a programme requires an openness about shortcomings
which may compromise the *demonstrating* purpose (especially where intended
for funders or external bodies). The purpose of *understanding*, as a research
activity to explain why phenomena occur in programmes, may be in conflict
with the *improving* function, where practical explanations may share little with
established language learning theory, and with the *judging* function, that is,
stating whether a particular practice is effective or not. Second, the *interaction*
and *collaboration* posited as central to learning from and through evaluation is
still novel in many programme contexts, where teachers work in isolation in
their classrooms, and the increasing demands (from bureaucratic, workload,
and technological sources) of classroom-based action leave little time for
teamwork. Third, it is unclear how to align learning through and learning from
evaluation, in particular in terms of benefits to teachers who engage actively in
evaluation. Where they focus on learning through, their learning may remain
invisible, with uncertain consequences for career development (Kiely, 2001).
Where they focus on learning from, the goal is dissemination through publi-
cations which may detract from dialogue and action stages in the programme
context. Enhancing learning from evaluation involves engagement with these
tensions, so that language programmes and the lives of programme partici-
pants are enriched rather than overwhelmed by the complex array of factors
which contribute to them.

In this paper I argue for a new synthesis in our conception of programme
evaluation which: (a) retains the focus on understanding and demonstrating

instructional effectiveness; (b) accounts for learning in both social and cognitive terms; and (c) engages with the features of context which have in the past been viewed as contextual constraints not relevant to the core task of the evaluation (innovation, teachers at work, and the quality of the learning experience as perceived by students). Accordingly, I first revisit the SLA-informed methods evaluations which dominated early language programme evaluation to illustrate the problem with narrow conceptions of language programmes. I then relate the three features of context which have particular relevance to the learning agenda within a materials evaluation in an English for Academic Purposes (EAP) programme.

It is important to note that the approach to evaluation set out here does not involve a diminution in any way of three established dimensions of good programme evaluation practice: assessment, research, and teacher development. Summative and formative *assessments* are important components of programme evaluation. The former constitutes a statement about learning gain within the programme, and an easily accessible data set for teachers and other evaluators. The latter represents a classroom-based process of monitoring and supporting learning (Rea-Dickins, 2001; Norris, 2006), reflecting in policy terms (what teachers are required to do) and in practice the ways teachers act to progress learning. *Research*, whether conventional theory-building, more situated action research (Burns, 2005), or exploratory practice (Allwright, 2003), can complement enquiry for programme evaluation, afford novel perspectives on phenomena within programmes, and constitute a source of inspiration, personal fulfilment, and career development for teachers. Skilled researchers are also in a position to contribute to data management and analysis procedures in evaluations. *Teacher development*, for example within a reflective practice framework (Richards & Lockhart, 1994; Freeman & Richards, 1996; Richards & Farrell, 2005), can constitute a strand of programme development which both parallels and complements programme evaluation. These activities, then, can contribute to the implementation of programme evaluation as conceived by Rea-Dickins, Pennington, and Norris above, even though they do not alone reflect or replace the comprehensive nature of evaluation.

I Three contexts of learning from evaluation

There are three main contexts of learning from language programme evaluation: (a) research, that is, theory building; (b) policy development; and (c) professional practice, that is, classroom or curriculum development. The contribution to theory building involves establishing generalizable findings about second language learning and teaching. Early evaluations took such theory building as their focus: from the 1960s a range of language programme evaluations sought to establish the superiority of a given language learning theory (see Beretta, 1992 for a review). Studies such as the Colorado

(Scherer & Wertheimer, 1964) and Pennsylvania (Smith, 1970) Projects in the USA, Primary French in the UK (Burstall *et al.*, 1974), the Communicational Teaching Project in Bangalore (Beretta & Davies, 1985; Prabhu, 1987), and the Hong Kong Expatriate English Language Teachers Scheme (Kiely & Rea-Dickins, 2005) had a research focus in two ways. First, they were designed to validate or inform second language learning theories; the Bangalore evaluation, for example, examined the hypothesis that instruction was more effective where a particular type of communicative task-based instruction developed by psycholinguist N. S. Prabhu was used. Second, these evaluations were designed as experimental studies with a focus on measurable and comparable learning outcomes, thereby seeking to provide rigorous evidence regarding the effectiveness of an instructional approach.

In addition to their theory building purpose, these studies had a policy development dimension. They were carried out in educational environments with the purpose of identifying instructional methods which could then be established as a basis for practice in classrooms, materials design, and teacher education. This shared enterprise in language programme evaluation between language learning researchers in the academy and policy makers responsible for foreign language curricula had its heyday from the 1960s to the 1980s. However, recent studies demonstrate how the research dimension of evaluation still shapes views of improving language teaching. For example, Klapper and Rees (2003) compared the impact of teaching based on Focus of Form (FonF) and Focus on Forms (FonFs) in a university German foreign language programme. This longitudinal study used mainly test data from the various stages of the four year programme (including one year of residence abroad in Germany) to compare the effectiveness of these instructional strategies. While there are valuable findings for language teaching researchers here, there is little to support a decision to opt for either FonF or FonFs as a broad instructional policy. In large part, this gap is caused by the lack of process data: the evaluation produced little information on the learning activities and experiences of the students over the four years, and equally little on the actual classroom activities introduced by the teachers. Such research-oriented evaluations address important current issues in methods of instruction. Teachers and policy-makers (such as managers, inspectors, and teacher educators) are naturally interested in arguments and evidence of effectiveness related to such classroom options. However the realities of these studies show how challenging it is to derive lessons for language teaching policy-making from theory-based experimental evaluations.

Another source of policy formation comes from the management and inspection tradition in education. Here, the instructional strategies teachers use are shaped by guidance from the curriculum authority, such as a ministry of education, or a mandate from an accrediting body, such as the British Association of Lecturers in English for Academic Purposes (BALEAP). The learning here is from professional expertise accumulated through experience

of teaching and observing classrooms. There is a focus on practicality and on establishing a threshold of quality which can serve as a guarantee for prospective students and clients. For example, the accreditation requirements of BALEAP set out in Table 1 illustrate how an external body can shape teaching and learning activities within a programme. Key principles relate to group size and composition, student-teacher relationships, and the role of learning strategies and autonomy. While these principles may reflect developments in language learning theory (for example, references to learning strategies and autonomy), they largely issue from the professional and practical lessons accumulated by a community over time.

The BALEAP accreditation requirements underpin a statement about quality for prospective students and their sponsors. This quality is maintained through a series of evaluations in the inspection tradition, 'Scrutiny on behalf of the relevant authority' as Norris (1998, p. 215) puts it. There are three limitations with an exclusive reliance on evaluations within the inspection tradition as a means of understanding and improving language programmes. First, the focus is on a minimum standard rather than excellence. Second, since the principles are further interpreted for actual practice, and as cautious managers monitor teachers' close adherence to the guidelines, they may lessen the creativity of individual teachers and limit opportunities for beneficial innovations. Third, the monitoring and inspection processes focus on the stated criteria and documentary evidence, thus emphasizing compliance rather than situated and creative aspects of teaching quality (and the capacity of evaluation to illuminate the same).

Both of the learning processes outlined above – understanding through research, and shaping good practice through mandates and quality control frameworks – have the effect of marginalizing the individual teacher and the contribution she makes to improving programmes. These approaches require

Table 1 Teaching and learning accreditation requirements

BALEAP teaching and learning requirements
5.5.1 The class size will not normally be more than twelve students and under no circumstances should it exceed sixteen. However, for certain activities classes may be combined, e.g. when giving students experience of lectures.
5.5.2 Students will be grouped into classes on a principled basis, according to factors such as proficiency level, gender, subject and nationality mix.
5.5.3 Students will be taught by a minimum of two main course teachers and not more than six different teachers during the week.
5.5.4 Adequate provision will be made for regular timetabled individual consultations or tutorials
5.5.5 Students will be encouraged to improve their own learning strategies and to develop as independent learners.

Source: http://www.baleap.org.uk/accreditation/index.aspx

that the individual teacher interpret principles or theories for her specific instructional context, while at the same time suggesting that it is not the teacher, but the researcher or the policy-maker who is the one to make such a specification. The challenge of learning from evaluation is how to integrate theory, policy, and the experience of teachers so that they may jointly shape understanding and developmental action in programmes.

One such evaluation study in the public domain (no doubt there are many such accounts in programmes and schools which have not been written up) is Towell and Tomlinson (1999). They provided a 10-year longitudinal account of a French as a foreign language programme in a British university. The authors, who are teachers and programme leaders as well as researchers, described the development and implementation of a task-based learning approach grounded in a second language acquisition perspective on language learning. They combined test data, students' views from questionnaire studies, and reflective comments from tutors to provide an account rich in opportunities for learning and programme development. One comment in their conclusion emphasizes the limitations of the study, and reveals their epistemological position: 'Curriculum design, application, evaluation and enhancement is a slow process, and subject to a number of extraneous influences which make it impossible to measure with totally scientific precision' (p. 25).

Clearly, this view privileges the psychometric (measurable) dimension of the programme, and marginalizes the social, cultural, and historical aspects – *extraneous influences* – which are likely to (a) shape the interpretation and implementation of the various curriculum initiatives, and (b) determine the engagement of teachers and students which probably account for the measured outcomes. The task of programme evaluation thus involves complementing this psychometric approach with a more qualitative view. As Holliday (2002) observes, 'Although the deeper social world is not quantifiable, its real nature can be established with sufficient weight of description' (p. 20).

An approach to programme evaluation which engages with the social world of language programmes, and achieves sufficient weight of description, is a methodological requirement to ensure balanced attention to the different contexts of learning – theory, policy, and practice – and to contribute in a dynamic and mutually informing way to the development of these different domains. The next section examines three aspects of programmes which have received limited attention in the past, but which afford additional (important) opportunities for learning from evaluation.

II Features of context in language programme evaluation

Innovation, teachers at work, and the quality of the student learning experience are aspects of programmes which link their theoretical constructs with their implementation as learning opportunities. The discussion in this section draws on developments outside language programme evaluation which can

enhance our understanding of how programmes are shaped. In Section III below, these developments are further related to the evaluation of learning materials in an EAP programme.

1 Innovation

Markee (1993) defines innovation in our field as 'proposals for qualitative change in pedagogical materials, approaches and values that are perceived as new by individuals who comprise a formal (language) education system' (p. 60). Innovations have often been the motivation for and focus of programme evaluations, even when this dimension of the programme is not emphasized. When an innovation is introduced, there is inevitably a settling in period when the practice may not be at its most effective. A series of adjustments and a process of learning about the unforeseen consequences of the innovation may be required before the potential of the innovation becomes realised in the practice. Chambers (1995) describes this process as a series of small innovations which follow a major innovation as part of the process of achieving fitness for purpose. Pennington (2004) describes a three-phase process by which: 'the innovation becomes naturalised to the context:

Time for an adjustment to occur between innovation and context;
Familiarity of more people to the context;
Extension of the innovation to more realms and functions' (p. 12).

There are two possible responses for evaluation here. First, an appropriate time for an evaluation might be after this naturalization period; when the innovation has become part of routine, skilled practice, its potential for the development of that programme can be understood. Van den Branden (2006), commenting on ten years of developing task-based teaching, emphasizes the complex process of implementing educational innovations: 'Many decisions have to be taken on the basis of expert intuition, literature studies, and lively discussions between the many parties involved, rather than on the basis of empirical evidence' (p. 16).

Second, the innovative programme and evaluation start together, and there is an explicit focus on the innovation process within the programme. Thus, problems and issues which arise can be understood primarily as an aspect of the novelty, or primarily as a feature of the planned action. An example of such a focus on innovation is the evaluation of the Primary Modern Languages Project (PMLP) in Ireland (Harris & O'Leary, 2002; Kiely & Rea-Dickins, 2005). The evaluation of the pilot study explicitly set out to document the programme as an innovation, attending to level and development of teacher skills, and the way schools coped:

> Because our evaluation work began with the project [PMLP] itself – unlike the situation in the case of other corresponding evaluations of modern languages in primary school – we were able to examine the start-up phase in some detail rather than just the end result

after two years. In particular we have been able to document teachers' early experience during the start up phase – something which other evaluations have not been able to do. (Harris & O'Leary, 2002, p. 21)

This evaluation thus was a rare opportunity to understand a programme as an innovation, and this understanding may have contributed to the recommendation in the evaluation report to proceed with differently shaped projects, focussing on clustering, continuity in learning, language awareness programmes, and intensive teacher development to ensure that provision of foreign languages in primary schools matched the needs and resources of participating schools and teachers.

Innovation is of course, part of all programmes. The focus of the discussion above is planned or policy-led innovations. In addition to these, there is innovative activity in all parts and at all stages of language programmes. Programmes introduce new resources such as computer assisted language learning or new course books; teachers try out new classroom tasks and assessment formats; students undertake activities and performances new to them. The drive to try different ways of doing things may be a fundamental aspect of the human condition, may be a strategy for relieving tedium and mechanical routine, or may simply reflect that each implementation of a programme is different. Evaluations should document both the nature and provenance of such variation, the demands in terms of teachers' time and effort, and also the capacity of a programme team to adapt to and maximize the learning potential of emerging opportunities. Such accounts would enhance policy and practice in the fields of task and materials development and teacher education, and thus enhance the learning potential of programme evaluation.

2 Programme evaluation and teachers at work

In major language programme evaluations of the past, teaching was often viewed as the delivery of a particular programme. The emphasis was on cognitive dimensions of learning and the effectiveness of a particular instructional strategy in achieving that. The teacher constituted a possible threat to blueprint fidelity. In recent years there has been a renewed interest in the contribution of teachers to language programmes, and three studies in particular suggest that their interpretations and decisions cannot be sidelined in any enterprise to understand the potential of the programme as a learning experience. Breen *et al*. (2001) developed the notion of 'teaching principles': 'reasons teachers give for particular techniques that they adopted during language lessons [which] revealed a set of guiding principles that appeared to be shared across the group' (p. 472).

Their analysis showed that lessons do not have either an exclusive content (i.e. language) or student focus, but rather a dynamic balancing of these two dimensions of the teaching task. This study complements analyses in Woods (1996) and Kiely (2001) which showed that teachers' methodological

decision making – *how* to teach – is shaped by the unfolding classroom inter-action. Richards (2006) explored one way in which such decision-making can be understood through an identity-oriented analysis of classroom interaction. He utilized a framework from Zimmerman (1998) which has three aspects of identity:

> *Discourse identity* which relates to the sequential development of the talk as participants engage as 'current speaker', 'listener', 'questioner', etc.
>
> *Situated identity* which refers to alignment to the identities of 'teacher' and 'student'.
>
> *Transportable identity* which refers to identities from the social world beyond the class-room which can be introduced – claimed for oneself or assigned to another – into the classroom discourse to add a dimension to the situated identity above, and thus alter the character of classroom talk. [Adapted from Richards (2006, p. 60)]

Richards's analysis is important in two ways for our understanding of lan-guage teaching, particularly in the context of language programme evaluation. First, teachers (and indeed students) can manage identities, for example, by introducing transportable identities in the classroom discourse and the lan-guage programme as a whole which make that classroom a uniquely engaging (or disengaging) set of learning opportunities. Second, teachers are likely to do this identity work to add value to the programme, that is, to enliven, to engage, to defy the tedium of the predicted and the planned.

The dynamic nature of Breen *et al*.'s pedagogic principles, and the identity work that Richards describes, might be viewed as attempts to relieve what Rampton (1999) refers to as 'the dull referentiality of school' (p. 335). Rampton developed a critique of conventional second language teaching based on Bernstein's notion of ritual. The ritualistic, routine aspects of a lan-guage programme, such as grammar drills, or teachers correcting students' work, may be meaningful because participants see them as valid activities and invest time and effort in them. Changes may therefore be viewed as dimin-ution of opportunities for learning. The important point for this discussion of evaluation and learning is that variation and change in the classroom may be difficult to achieve consensually, effectively, and quickly. The Rampton analysis reminds us of the limited scope teachers may have to shape the pro-gramme process, and to develop the creative learning spaces which charac-terize language learning opportunities in classrooms and programmes.

From these studies we can see teachers' work less as uncomplicated implementation of a pedagogic script in the form of a syllabus, course book, or scheme of work – the traditional evaluator's dream – and more as an indi-vidualized struggle to identify and develop the small spaces for pedagogic and interactional variation and unpredictability.

3 The quality of the learning experience of students

A programme has to work at the level of experience. The learning experience that is satisfying in a holistic way has the potential to engage, motivate,

generate effort, and lead to desired outcomes. Studies such as Towell and Tomlinson (1999) illustrate the importance of a sustained positive view of the classroom experience. This is an aspect of evaluation which is widely implemented: programmes typically end with a questionnaire where students indicate their satisfaction with the programme. The argument here relates to the extent to which this data is used to understand the programme as a learning experience. The motivation for the student satisfaction questionnaire is very often to demonstrate to institutions and accrediting bodies which require such practice as a quality assurance mechanism, or as evidence of attending to the perspectives of learners (Crabbe, 2003; Mackay, 1994; Mackay, *et al*., 1995). A focus in evaluation on the quality of the learning experience relates to learning in two ways. First, the general impression captured through rating scales and tick-box questionnaires can accumulate over time, and through careful analysis and occasional augmentation by interview or group discussion studies, evaluations can provide an invaluable account of student concerns, wants, expectations, and investments. Second, participation in this form of evaluation can enhance student learning in two ways: achievement in terms of the language learning goals of the programme, and enhanced understanding of the process of learning, particularly in relation to individual style, capacity and dispositions. Legutke and Thomas's (1991) discussion of the role of student evaluation reflects these two goals:

> Engaging learners in communicative encounters, especially if their aim is to explore emotional content and experiences, can become too bound up in itself unless this activity also reaches an evaluation stage. Trying to understand what has happened while undertaking a particular task, why it was suggested by the teacher, and contributing actively to the evaluation of learning arrangements, sequences, resources and input materials by means of reflection and meta-communicative discourse – all these are considered indispensable learner activities in ELT. (p. 65)

A focus on the quality of the learning experience of students in programme evaluations complements the discussion above about understanding innovations and innovation capacity, and engaging with the situated nature of teachers' at work on programmes. A satisfying learning experience sustains motivation, develops a learning identity and trajectory, and nurtures investment in learning activity.

III EAP materials evaluation

In this section I illustrate the preceding ideas in an evaluation of an EAP programme in a British university. The stated purposes of evaluation in this institution were accountability (demonstrating to local stakeholders and external bodies that the programme was well planned, staffed, and taught) and development (facilitating ongoing improvement of the EAP provision for current and future students). The evaluation process was devolved to programmes

and tutors: each EAP tutor facilitated a mid-course structured discussion, administered an end-of-course questionnaire, and submitted a brief report of issues raised and actions taken to the EAP coordinator. She in turn summarized issues from all such evaluations for the institution's academic committees which had responsibility for monitoring academic quality. The focus of the discussion here is the evaluation of a particular type of EAP learning materials, and the data are from two studies: the programme evaluation, and an ethnographic research study into the impact of evaluation processes. Table 2 summarizes the key features of the two studies.

A feature of this evaluation was the limited data which emerged from it, just a brief report on issues and actions. More valuable was the process in the classroom. The discussion between teachers and students shaped teaching strategies in the remaining weeks of the programme (Kiely, 2001), and provided opportunities for critical thinking on the part of students (Kiely, 2004). There was learning from and through the evaluation, but this was not the whole story. The ethnographic study also illuminated learning opportunities not taken, and one of these – learning materials – is the focus of analysis here. The critical issue in focus was the perceived and actual benefit of lists of discourse markers for the development of academic writing skills. I have selected one such list here (there were six in all in the first six weeks of the 12-week programme) used in Week 3. The data on the use of this handout and the perspectives of the teacher (Anna) and one student (Laure) profile clearly some issues of effectiveness and show how the aspects outlined above – innovation, teachers at work, and quality of the student learning experience – afford a fuller explanation of the phenomena observed and the forces shaping these. Such an explanation is a useful starting point for enhancing learning from programme evaluation.

The materials presented a list of discourse markers fulfilling a particular textual function – *addition*; *transition*; *contrast*, and so forth. The list for

Table 2 Programme evaluation and ethnographic study

	Programme evaluation	Ethnographic study
Purpose	Accountability Development	Research into impact of evaluation processes
Participants	Teacher; Students (17)	Teacher; Students (17)
		EAP coordinator Ethnographer
Data	Structured discussion End of course questionnaire Written summary of issues and actions	Fieldnotes Learning materials Assessment documentation Interviews (with students, teacher, and coordinator)

Table 3 Transition handout

Transition – helps you move on to a new stage of the argument you are developing
Now With reference to With regard to With respect to As for As to

Transition is set out in Table 3. In class, the teacher explained what these meant and provided oral examples of how they linked ideas:

Anna: I will give more words to structure your writing now – connectives, or semantic mark-ers, whatever you want to call them.

Anna puts on OHP and reads from it: Transition. *Then tells students that there is a handout to follow. Then Anna talks to the topic:* When you want to change course, these are linking words and phrases you can use.

About six students are taking notes at this stage – have they missed the announcement that there is a handout? Or are the words and phrases listed just very 'takedownable'?

Anna: A problem some students have is they do not have smooth transitions, perhaps not para-graph divisions. Also a lot of this is a game: you have to find the key and then reproduce it on demand. Think of a lecturer marking this Sunday lunchtime (*Anna is smiling*), and it is a long rambling thing, lecturers get irritated because it is so disorganised. These words and phrases can help the structure.

<div align="right">Classroom notes</div>

(Unmarked font: actual words; *Italics: observation notes and interpretations*)

The students were very positive about these lists, both in the group discussion which was part of the programme evaluation activity (see Table 2) and in interview. The comments in Table 4 reflect how one student, Laure, felt they would be effective in developing her writing skills:

As it turned out, however, Laure did not benefit from the assumed potential of these materials. When it came to writing an essay towards the end of the programme, she found the lists of discourse markers were of limited value:

Laure: The problem is that I wrote my essay without any notes, only helping me with the books and the dictionary, that I wanted to put the definition of education. […] When I typed it, I took the piece of paper about the link words to change, to really make, for example, as an introduction, I will say, or we, I think I used we, we will say, and as a conclusion, to really make a difference between introduction, conclusion, in the first part we will speak, firstly, secondly, or I said at first or at the beginning, and then I said moreover, and finally, and the same in the second part, I tried to organise my work, I tried, but I don't know if it is really, really good

<div align="right">(Interview with Laure)</div>

Table 4 Benefits of the word list for Laure

Laure's view (interview)	Benefits
I think it is really great because, in fact, when we are here we have to write a lot of essays, and the problem that sometimes you don't put the correct link word, and it is not really easy after that for the teacher to see the full meaning of the plan and everything. [...] I think it will help me to improve my writing, how to write more properly an essay	Useful for structuring essays generally
This one (pointing to Transition list) was really useful, we learnt the *transition*, all the link words, it is useful for the essay, I will use it for the next task	Relevant to a specific writing task
I will read them again when I will do my essay to try to introduce some words from this piece of paper to my work	Useful when preparing to write
I keep them in a, a file, [...] I bring them with me to every lecture	Useful as a portable resource

Laure's experience was not a surprise to Anna, the teacher who was aware of the limitations of developing writing skills through word lists:

Anna: [talking about handouts on EAP discourse markers] One of the bits I want to develop – at some point I am going to do it, I don't know if I am going to have time to do it this module, but where I am using those words, you know, connectives etc., what I want to do is have examples from academic texts where those things are used, to give them, I've done bits and pieces before, I would like to do it more systematically. So an example of *listing* from a text, from any academic text, to show how skilled academic writers in English might use it.

(Interview with Anna)

Anna's intention here is revealing. She is aware of the limited value of these materials which are so popular with students. She can see how value might be added, by situating the discourse markers in a text context. The time factor was the key reason why she had not undertaken this development of her EAP materials. The practice appears to be sustained by two additional factors. First, the materials are appreciated by students, in part because the list each week was a portable product they could take away from the classroom, and in part because they appeared to relate directly and practically to the challenge of academic writing in English. Thus, the students' response to the classroom process sustained the materials in list form, constructing a shared investment in the activity as a fruitful and valued opportunity for learning. Second, this synthetic, decontextualized approach to teaching

aspects of discourse worked well in terms of the teacher's instructor role. It enabled her to lead, to focus, to pronounce authoritatively in a manner which can be aligned to institutionally assigned roles and expected rituals, such as the students taking notes, even when a handout is available. The alternative pedagogy based on discourse markers embedded in actual EAP text, thus, might have been seen as constituting a time-consuming innovation for the teacher, which ran the risk of student dissatisfaction: it would be both difficult to organise and challenging to implement. The task of programme evaluation here is to map these considerations, so that the participants, particularly the teacher and the host institution, understand the complex interrelationships and can develop the materials and thus learning opportunities which take into account the three factors above.

Innovation: Anna's own evaluation of these materials is characterized by the need to innovate, because she knows they are not really effective as decontextualized lists. The stated constraint is time, a value-free factor which often obscures other reasons. In this case, the programme context may be one where there is limited capacity to evaluate. Students on short programmes (12 weeks) work largely with conceptions of language skills development from previous courses, and teachers like Anna may judge that this is not long enough to innovate effectively, that is, to develop a new student perspective on classroom activities. For a different practice to become naturalized (Pennington, 2004), careful mapping of the space between the old and the new is required, and students need to be guided through this space.

Teachers at work: The classroom episode above illustrates the complexity of the teacher's identity work (Richards, 2006). Anna starts off in the default or situated identity as teacher. She then shifts, and affiliating with the students' challenges as writer, refers to the 'game' and finding the 'key'. Then, a transportable identity as reader and marker is drawn on when she refers to the 'irritated lecturer' marking this 'Sunday lunchtime' (the smile indicating perhaps that this is indeed her own transported identity rather than a reference to some mythical other). The overall purpose of the discourse here is to establish a pedagogic case for the materials, a case which, as indicated by Anna in interview, is somewhat tenuous. The case Anna is making is the rationale for the specific *Transition* list, but also for the programme as a whole. Anna wants students to feel that the programme is worth investing in, that they can accept it as a viable vehicle for achieving their own learning goals as well as the programme objectives.

In terms of Breen *et al.*'s (2001) principles, Anna shows a 'concern with how to use the classroom and its human and material resources to optimise learning' (p. 484), but this could be considered overly respectful of the ritual of the teacher authoritatively taking students through a list. The activity conforms to expectations of the classroom and to Laure's own view of her role as student (see Table 4), and also reflects what Rampton (1999) describes as the power of ritual in language classrooms. Anna is thus working with a complex

set of factors which shape the classroom, and which leave her with limited capacity for ameliorative action.

Quality of the learning experience for students: Clearly, the positive views of these participants need to be maintained. The beliefs and investment practices of Laure reflect a positive programme experience. However, the link to outcomes, that is, evidence of learning gain in her writing skills, resulting from materials designed to this end, also needs to be considered. Students like Laure may not be able to do such linking; for them, the outcomes are the future, belief is blind, and investments are acts of faith. The teachers and programme managers can know from evaluation and reflection whether the assumed learning opportunities of this particular activity are viable, and they can use this knowledge to orient student thinking and learning strategies.

The examination of this evaluation shows how the programme is a complex entity, with both moving parts and fixed angles. It shows the influence of biography – what teachers and students bring to programmes – and how this can be both a stimulus for and brake on change and development. The short duration, a matter of calendar convenience rather than a principle of curriculum design, may be a major factor in cost-benefit analyses of change and innovation. The teachers' leadership role in providing materials, and her responsibility in leading the classroom process and organizing learning opportunities, can facilitate a shared dynamism and enthusiasm in the classroom, or lead to instructional routines which are positively received but may be limited in facilitating learning gain.

The analysis presented here is based in part on the evaluation, and in part on the ethnographic research study of evaluation processes. The latter, based largely on a structured discussion in the classroom, was effective in terms of *judging*, *demonstrating* and *improving* (Norris, 2006). However, the research study illustrates shortcomings in terms of *understanding*, and consequently in improving the programme through transformation of learning materials and thus teacher and student roles. Tensions in the evaluation procedure can be identified at three levels. First, the teacher was both the leader and subject of the evaluation, and management of the process involved a justification of her practice. Second, at the institutional level, the policy of evaluation for accountability and development proved problematic: it privileged student preferences over the teacher's account of her work, and the defence of the status quo over innovation and change. And third, at the level of prevailing theory, the approach to learning involved teachers and students in a provider-consumer relationship. Within this framework, lists of words as learning materials worked satisfactorily, even though their shortcomings for learning were apparent.

IV Conclusion

The title of this paper, 'small answers to the big question', relates to the programme effectiveness question, which is at the core of all our conceptions of

evaluation. In the past, the search for effectiveness has been through grand theories, branded methods, and passing trends. The search for the secret of effectiveness has often focussed on the nature of the programme inputs and the measurement of outcomes. This article proposes a more distributed focus. As well as inputs and outcomes, there is a need to examine the interactions and the factors which shape input use and impact. Any programme can be effective in promoting language learning. The key issues are how it has been made to work, and which factors and events have shaped success. Certainly, there will be many answers to this question, phenomena minor in themselves but connecting with a range of other features of the programme to constitute a complex set of experiences, with more or less potential for learning.

Evaluation then becomes a set of strategies to document and understand the programme. It involves research activity (conventional studies or action research by which teachers learn about and transform aspects of their practice) and assessment data (conventional measures of outcomes). In addition to these, evaluation has to engage with the social, cultural, and historical identity of the programme, as a product of the institution, as a phase in the biographies of participants, and as a context of personal investments of individual stakeholders. Learning from and learning through this form of evaluation may facilitate two-way traffic in the knowledge-building enterprise: theories and concepts from outside language programmes becoming available as resources for development, and constructs, understandings, and explanations from practice informing research and theory-building domains in applied linguistics and language education. A key outcome of this traffic will be policies which shape programmes, teacher education, and materials, and which guide the work of teachers through quality management frameworks that are better informed by both theory and practice.

Acknowledgements

V References

Allwright, D. (2003). Exploratory practice: Rethinking practitioner research in language teaching. *Language Teaching Research, 7*(2), 113–141.

Beretta, A. (1992). Evaluation of language education: An overview. In J. C. Alderson & A. Beretta, (Eds.), *Evaluating second language education* (pp. 5–24). Cambridge: Cambridge University Press.

Beretta, A., & Davies, A. (1985). Evaluation of the Bangalore Project. *English Language Teaching Journal, 39*(2), 121–127.

Breen, M. P., Hird, B., Milton, M., Oliver, R., & Thwaite, A. (2001). Making sense of language teaching: Teachers' principles and classroom practices. *Applied Linguistics, 22*(4), 470–501.

Burns, A. (2005). Action research: An evolving paradigm? *Language Teaching, 38*(2), 57–74.

Burstall, C., Jamieson, M., Cohen, S., & Hargreaves, M. (1974). *Primary French in the balance*. Slough: National Foundation for Educational Research.

Chambers, F. (1995). Cyclic innovation in ELT: implications for implementation. In O'Dwyer, J. (Ed.) (1996), *Quality Management and the Management of Change: Papers presented at an International Conference on Management in English Language Teaching* (pp. 67–76). Ankara, Turkey: Bilkent University School of English Language.

Crabbe, D. (2003). The quality of language learning opportunities. *TESOL Quarterly, 37*(1), 9–34.

Freeman, D., & Richards, J. C. (Eds.), (1996). *Teacher learning in language teaching*. Cambridge: Cambridge University Press.

Harris, J., & O'Leary, D. (2002) *National survey of principals of schools involved in the Modern Languages in Primary Schools Initiative in Ireland*. Dublin: Institiúid Teangeolaíochta Éireann.

Holliday, A. (2002). *Doing and writing qualitative research*. London: SAGE Publications.

Kiely, R. (2001). Classroom evaluation – values, interests and teacher development. *Language Teaching Research, 5*(3), 241–261.

Kiely, R. (2004). Learning to critique in EAP. *Journal of English for Academic Purposes, 3*(3), 211–227

Kiely, R., & Rea-Dickins, P. (2005). *Programme evaluation in language education*. Basingstoke: Palgrave.

Klapper, J., & Rees, J. (2003). Reviewing the case for explicit grammar instruction in the university foreign language learning context. *Language Teaching Research, 7*(3), 285–314.

Legutke, M., & Thomas, H. (1991). *Process and experience in the language classroom*. Harlow: Longman.

Mackay, R. (1994). Undertaking ESL/EFL programme review for accountability and improvement. *English Language Teaching Journal, 48*(2), 142–149.

Mackay, R., Wellesley, S., & Bazergan, E. (1995). Participatory evaluation. *English Language Teaching Journal, 49*(4), 308–317.

Markee, N. (1993). The diffusion of innovation in language teaching. *Annual Review of Applied Linguistics, 13*, 229–243.

Norris, J. M. (2006). The why (and how) of student learning outcomes assessment in college FL education. *Modern Language Journal, 90*(4), 576–583.

Norris, N. (1998). Curriculum evaluation revisited. *Cambridge Journal of Education, 28*(2), 207–220.

Pennington, M. (1998). Designing language programme evaluations for specific purposes. In E. Li & James, G. (Eds.), *Testing and evaluation in second language*

education (pp. 199–213). Hong Kong: Hong Kong University of Science and Technology.

Pennington, M. (2004). Cycles of innovation in the adoption of information technology: A view for language teaching. *Computer Assisted Language Learning, 17*(2), 7–33.

Prabhu, N. S. (1987). *Second language pedagogy*. Oxford: Oxford University Press.

Rampton, B. (1999). Dichotomies, difference, and ritual in second language learning and teaching. *Applied Linguistics, 20*(3), 316–340.

Rea-Dickins, P. (1994). Evaluation and English language teaching. *Language Teaching, 27*(2), 71–92.

Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes in classroom assessment. *Language Testing, 18*(4), 428–462.

Richards, K. (2006). Being the teacher: Identity and classroom conversation. *Applied Linguistics, 27*(1), 51–77.

Richards, J. C., & Farrell, T. (2005). *Professional development for language teachers*. Cambridge: Cambridge University Press

Richards, J. C., & Lockhart, C. (1994). *Reflective teaching in second language classrooms*. Cambridge: Cambridge University Press.

Scherer, G. A. C., & Wertheimer, M. (1964). *A psycholinguistic experiment in foreign language teaching*. New York: McGraw-Hill.

Smith, P. D. (1970). *A comparison of the cognitive and audio-lingual approaches to foreign language instruction: The Pennsylvania Foreign Language Project*. Philadelphia, PA: The Center for Curriculum Development.

Towell, R., & Tomlinson, P. (1999). Language curriculum development research at university level. *Language Teaching Research, 3*(1), 1–32.

Van den Branden, K. (2006). *Task-based language education: From theory to practice*. Cambridge: Cambridge University Press

Woods, D. (1996). *Teacher cognition in language teaching*. Cambridge: Cambridge University Press.

Zimmerman, D. H. (1998). Discoursal identities and social identities. In C. Antaki & S. Widdicombe (Eds.), *Identities in talk* (pp. 87–106). London: SAGE Publications.