The empirical evaluation of language teaching materials

Rod Ellis

This article distinguishes two types of materials evaluation: a predictive evaluation designed to make a decision regarding what materials to use, and a retrospective evaluation designed to examine materials that have actually been used. Retrospective evaluations can be impressionistic or empirical. It is suggested that one way in which teachers can conduct empirical evaluations is by investigating specific teaching tasks. A procedure for conducting a task evaluation is described. Finally, it is suggested that task evaluations constitute a kind of action research that can contribute to reflective practice in teaching.

Materials evaluation: an overview

Teachers are often faced with the task of choosing what teaching materials to use. In effect, they are required to carry out a *predictive evaluation* of the materials available to them in order to determine which are best suited to their purposes. Then, once they have used the materials, they may feel the need to undertake a further evaluation to determine whether the materials have 'worked' for them. This constitutes a *retrospective evaluation.*

Predictive evaluation

A brief review of the literature relating to materials evaluation reveals that, to date, the focus of attention has been more or less exclusively on predictive evaluation. There are two principal ways in which teachers can carry out this kind of evaluation. One is to rely on evaluations carried out by 'expert' reviewers. Journals like *ELT Journal* assist teachers in this respect by providing reviews of published coursebooks. In some cases (such as the Survey Reviews this journal provides from time to time), the reviewers identify specific criteria for evaluating materials. However, in reviews of individual coursebooks, the criteria often remain inexact and implicit.

Alternatively, teachers can carry out their own predictive evaluations. There are numerous checklists and guidelines available to help them do so (e.g. Cunningsworth 1984, Breen and Candlin 1987, Skierso 1991, McDonough and Shaw 1993). These instruments are generally organized in a manner that reflects the decision-making process which it is hypothesized teachers go through. Breen and Candlin (1987), for example, organize the questions in their checklist into two phases, the first of which enables teachers to address the overall 'usefulness' of the materials, while the second caters for 'a more searching analysis' based on the teacher's actual teaching situation. The idea behind these guides is to help teachers carry out a predictive evaluation *systematically.*

However, there are limits to how 'scientific' such an evaluation can be. As Sheldon (1988: 245) observes, 'it is clear that coursebook assessment is fundamentally a subjective, rule-of-thumb activity, and that no neat formula, grid or system will ever provide a definite yardstick'.

Retrospective
evaluation

This being so, the need to evaluate materials retrospectively takes on special importance. Such an evaluation provides the teacher with information which can be used to determine whether it is worthwhile using the materials again, which activities 'work' and which do not, and how to modify the materials to make them more effective for future use. A retrospective evaluation also serves as a means of 'testing' the validity of a predictive evaluation, and may point to ways in which the predictive instruments can be improved for future use.

Somewhat surprisingly, however, there are very few published accounts of retrospective evaluations of course materials, and very little information about how to conduct them. The bulk of the published literature on evaluation deals with programme or project evaluation (e.g. Alderson 1992, Weir and Roberts 1994, Lynch 1996). Such evaluations may incorporate materials evaluation but they are necessarily much broader in scope. Otherwise, the only other published work on the empirical evaluation of teaching materials is to be found in accounts of the trialling of new materials (e.g. Barnard and Randall 1995). The purpose of this article is to begin to address the question of how retrospective evaluations of materials can be carried out.

Evaluating course
materials
retrospectively

Teachers can perform a retrospective evaluation impressionistically or they can attempt to collect information in a more systematic manner (i.e. conduct an empirical evaluation). It is probably true to say that most teachers do carry out impressionistic evaluations of their teaching materials. That is, during the course they assess whether particular activities 'work' (usually with reference to the enthusiasm and degree of involvement manifested by the students), while at the end of the course they make summative judgements of the materials. Empirical evaluations are perhaps less common, if only because they are time-consuming. However, teachers report using students' journals and end-of-course questionnaires to judge the effectiveness of their teaching, including the materials they used.

One way in which an empirical evaluation can be made more manageable is through micro-evaluation. A macro-evaluation calls for an overall assessment of whether an entire set of materials has worked. To plan and collect the necessary information for such as empirical evaluation is a daunting prospect. In a micro-evaluation, however, the teacher selects one particular teaching task in which he or she has a special interest, and submits this to a detailed empirical evaluation. A series of micro-evaluations can provide the basis for a subsequent macro-evaluation. However, a micro-evaluation can also stand by itself and can serve as a practical and legitimate way of conducting an empirical evaluation of teaching materials.

Conducting a micro-evaluation of tasks

A micro-evaluation of teaching materials is perhaps best carried out in relation to 'task'. This term is now widely used in language teaching methodology (e.g. Prabhu 1987; Nunan 1989), often with very different meanings. Following Skehan (1996), a task is here viewed as 'an activity in which: meaning is primary; there is some sort of relationship to the real world; task completion has some priority; and the assessment of task performance is in terms of task outcome'. Thus, the information and opinion-gap activities common in communicative language teaching are 'tasks'.

Describing a task

A 'task' can be described in terms of its *objectives*; the *input* it provides for the students to work on (i.e. the verbal or non-verbal information supplied); the *conditions* under which the task is to be performed (e.g. whether in lockstep with the whole class or in small group work); the *procedures* the students need to carry out to complete the task (e.g. whether the students have the opportunity to plan prior to performing the task); and *outcomes* (i.e. what is achieved on completion of the task). The outcomes take the form of the product(s) the students will accomplish (e.g. drawing a map, a written paragraph, some kind of decision) and the processes that will be engaged in performing the task (e.g. negotiating meaning when some communication problem arises, correcting other students' errors, asking questions to extend a topic).

Evaluating a task

Evaluating a task involves a series of steps:

Step 1: Choosing a task to evaluate
Step 2: Describing the task
Step 3: Planning the evaluation
Step 4: Collecting the information for the evaluation
Step 5: Analysing the information
Step 6: Reaching conclusions and making recommendations
Step 7: Writing the report

Choosing a task to evaluate

Teachers might have a number of reasons for selecting a task to micro-evaluate. They may want to try out a new kind of task and be interested in discovering how effective this innovation is in their classrooms. On other occasions they may wish to choose a very familiar task to discover if it really works as well as they think it does. Or they may want to experiment with a task they have used before by making some change to the input, conditions, or procedures of a familiar task and decide to evaluate how this affects the outcomes of the task. For example, they may want to find out what effect giving learners the chance to plan prior to performing a task has on task outcomes.

Describing the task

A clear and explicit description of the task is a necessary preliminary to planning a micro-evaluation. As suggested above, a task can be described in terms of its objective(s), the input it provides, conditions, procedures, and the intended outcomes of the task.

Rod Ellis

Alderson (1992) suggests that planning a program evaluation involves working out answers to a number of questions concerning the purpose of the evaluation, audience, evaluator, content, method, and timing (see Figure 1). These questions also apply to the planning of a micro-evaluation. They should not be seen as mutually exclusive. For example, it is perfectly possible to carry out both an objectives model evaluation, where the purpose is to discover to what extent the task has accomplished the objectives set for it, and a development model evaluation, where the purpose is to find out how the task might be improved for future use, at one and the same time. The planning of the evaluation needs to be undertaken concurrently with the planning of the lesson. Only in this way can teachers be sure they will collect the necessary information to carry out the evaluation.

Figure 1:
Choices involved in
planning a task-
evaluation

Question	Choices
1 Purpose (Why?)	**a.** The task is evaluated to determine whether it has met its objectives (i.e. an objectives model evaluation).
	b. The task is evaluated with a view to discovering how it can be improved (i.e. a development model evaluation).
2 Audience (Who for?)	**a.** The teacher conducts the evaluation for him/herself.
	b. The teacher conducts the evaluation with a view to sharing the results with other teachers.
3 Evaluator (Who?)	**a.** The teacher teaching the task.
	b. An outsider (e.g. another teacher).
4 Content (What?)	**a.** Student-based evaluation (i.e. students' attitudes towards and opinions about the task are investigated).
	b. Response-based evaluation (i.e. the outcomes—products and processes—of the task are investigated).
	c. Learning-based evaluation (i.e. the extent to which any learning or skill/strategy development has occurred) is investigated.
5 Method (How?)	**a.** Using documentary information (e.g. a written product of the task).
	b. Using tests (e.g. a vocabulary test).
	c. Using observation (i.e. observing/recording the students while they perform the task).
	d. Self-report (e.g. a questionnaire to elicit the students' attitudes).
6 Timing (When?)	**a.** Before the task is taught (i.e. to collect baseline information).
	b. During the task (formative).
	c. After the task has been completed (summative): i) immediately after ii) after a period of time.

The decision on *what* to evaluate is at the heart of the planning process. Here three types of evaluation can be identified. In a student-based evaluation, the students' attitudes to the task are examined. The basis for such an evaluation is that a task can only be said to have worked if the students have found it enjoyable and/or useful. Evaluations conducted by means of short questionnaires or interviews with the

students are the easiest kind to carry out. Response-based evaluations require the teacher to examine the actual outcomes (both the products and processes of the task) to see whether they match the predicted outcomes. For example, if one of the purposes of the task is to stimulate active meaning negotiation on the part of the students, it will be necessary to observe them while they are performing the task to which they negotiate or, alternatively, to record their interactions for subsequent analysis in order to assess the extent to which they negotiate. Although response-based evaluations are time-consuming and quite demanding, they do provide valuable information regarding whether the task is achieving what it is intended to achieve. In learning-based evaluations, an attempt is made to determine whether the task has resulted in any new learning (e.g. of new vocabulary). This kind of evaluation is the most difficult to carry out because it generally requires the teacher to find out what the students know or can do before they perform the task *and* after they have performed it. Also, it may be difficult to measure the learning that has resulted from performing a single task. Most evaluations, therefore, will probably be student-based or response-based.

Collecting the information

.

As Figure 1 shows, the information needed to evaluate a task can be collected before, during, or after the teaching of the task. It may be useful for the evaluator to draw up a record sheet showing the various stages of the lesson, what types of data were collected, and when they were collected in relation to the stages of the lesson. This sheet can be organized into columns with the left-hand column showing the various stages of the lesson and the right-hand column indicating how and when information for the evaluation is to be collected.

Analysing the information

Two ways of analysing the data are possible. One involves quantification of the information, which can then be presented in the form of tables. The other is qualitative. Here the evaluator prepares a narrative description of the information, perhaps illustrated by quotations or protocols. In part, the method chosen will depend on the types of information which have been collected. Thus, test scores lend themselves to a quantitative analysis, while journal data is perhaps best handled qualitatively.

Reaching conclusions and making recommendations

It is useful to distinguish 'conclusions' and 'recommendations'. Conclusions are general statements about what has been discovered about the task from the analyses that have been performed. Recommendations are the evaluator's ideas regarding future actions. The conclusions need to be framed in relation to the purposes of the evaluation. Thus, in an objectives model evaluation, the conclusions need to state to what extent the objectives of the task have been met, while in a development model evaluation the conclusions need to indicate in what ways the task has worked or not worked, and how it can be improved.

Writing the report

Strictly speaking, it is not necessary to write a report of an evaluation unless the evaluator intends to share the conclusions and recommenda-

tions with others. However, by writing a report the teacher-evaluator is obliged to make explicit the procedures that have been followed in the evaluation and, thereby, is more likely to understand the strengths and limitations of the evaluation.

Conclusion · Materials have traditionally been evaluated predictively using checklists or questionnaires to determine their suitability for use in particular teaching contexts. There have been surprisingly few attempts to evaluate materials empirically, perhaps because a thorough evaluation of a complete set of materials is a daunting undertaking, which few teachers have the time to make. There is, however, an urgent need for the empirical evaluation of teaching materials. One way in which this might be made practical is through micro-evaluations of specific tasks. The purpose of this article has been to suggest how such micro-evaluations can be accomplished.

A micro-evaluation of a task can serve several purposes. It can show to what extent a task works for a particular group of learners. It can also reveal weaknesses in the design of a task, and thus ways in which it might be improved.

It can be argued that teachers have always engaged in evaluating the tasks they use and that the kind of micro-evaluation advocated here is, therefore, unnecessary. However, it can be counter argued that there is much to be gained by formalizing the procedures used to carry out micro-evaluations. First, the procedure that has been advocated in this article requires teachers to pay attention to evaluation as they plan lessons, as many educators advocate (e.g. Nunan 1988). Second, formalizing the procedure for evaluation forces teachers to go beyond impressionistic assessments by requiring them to determine exactly what it is they want to evaluate and how they can do it. Third, micro-evaluation serves as one way of conducting action research and, thereby, of encouraging the kind of reflection that is believed to contribute to teacher development (Richards and Lockhart 1994). In fact, teachers may find it easier to begin action research by identifying a task they would like to evaluate than by looking for a problem to solve, the usual way of getting started. Fourth, and perhaps most important, micro-evaluation serves as a form of professional empowerment. Clarke (1994: 23) has argued that teachers need 'to keep their own counsel regarding what works and does not work and to insist on an interpretation of events and ideas that includes . . . a validation of their own experiences in the classroom'. While this does not necessitate a commitment to systematic evaluation, it does assume a responsibility for ensuring that classroom events are interpreted as accurately and systematically as possible. Carefully planned materials evaluations, in the form of task evaluations, may provide a practical basis for achieving this.

Received May 1996

References

Alderson, J. 1992. 'Guidelines for the evaluation of language education' in J. Alderson and A. Beretta (eds.). *Evaluating Second Language Education.* Cambridge: Cambridge University Press.

Barnard, R. and **M. Randall.** 1995. 'Evaluating course materials: a contrastive study in textbook trialling'. *System* 23/3: 337–46.

Breen, M. and **C. Candlin.** 1987. 'Which materials? A consumer's and designer's guide' in L. Sheldon (ed.). *ELT Textbooks and Materials: Problems in Evaluation and Development.* ELT Documents 126. London: Modern English Publications.

Clarke, M. 1994. 'The dysfunctions of theory/practice discourse'. *TESOL Quarterly* 281: 9–26.

Cunningsworth, A. 1984. *Evaluating and Selecting ELT Materials.* London: Heinemann.

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

McDonough, J. and **C. Shaw.** 1993. *Materials and Methods in ELT.* Oxford: Blackwell.

Nunan, D. 1988. *The Learner-centred Curriculum.* Cambridge: Cambridge University Press.

Nunan, D. 1989. *Designing Tasks for the Communicative Classroom.* Cambridge: Cambridge University Press.

Prabhu, N. 1987. *Second Language Pedagogy.* Oxford: Oxford University Press.

Richards, J. and **C. Lockhart.** 1994. *Reflective Teaching in Second Language Classrooms.* Cambridge: Cambridge University Press.

Sheldon, L. 1988. 'Evaluating ELT textbooks and materials'. *ELT Journal* 42: 237–46.

Skehan, P. 1996. 'A framework for the implementation of task-based instruction'. *Applied Linguistics* 17/1: 38–62.

Skierso, A. 1991. 'Textbook selection and evaluation' in M. Celce-Murcia (ed.). *Teaching English as a Second or Foreign Language.* Boston: Heinle and Heinle.

Weir, C. and **J. Roberts.** 1994. *Evaluation in ELT.* Oxford: Blackwell.

The author

Rod Ellis is currently Professor of TESOL at Temple University, Philadelphia. He has worked in teacher education in Zambia, the United Kingdom, and Japan. He has published books on second language acquisition and teacher education and, in addition, a number of EFL/ESL textbooks.