

THE EVALUATION OF TEACHERS' GUIDES- DESIGN AND APPLICATION

Michael Hemsley

Abstract

This paper begins with a discussion of the functions of the teachers' guide (TG). This is followed by an examination of various issues in the design of TG evaluations, and several existing checklists of criteria for evaluation of TGs and ELT materials in general are introduced. The next section concerns itself with evaluation in the context of a private language school. Finally, a model for TG evaluation in this context is developed. It is hoped that readers will be able to adapt this model in order to carry out TG evaluations in their own contexts. Reference tables are provided in two appendices.

The teachers' guide, what it is, and what it does

The forms taken by TGs are many and varied, and any definition of something so heterogeneous would be incomplete. What we can do is attempt to list the functions of an ideal TG as do Cunningsworth and Kusel (1991:129). We may paraphrase them thus: The teachers' guide states the purpose of the associated teaching materials and describes the rationale behind them. The TG will encourage the development of teaching skills and assist the teacher to understand the course as a whole. It will provide guidance on how to use the material and the linguistic and cultural information required for its effective use.

To these five functions we could add that a good TG will also help teachers develop towards an eventual position of self-reliance and independence of such explicit guidance.

Servants and masters

The view of TGs outlined above raises a general question: If the guide is the master, does the teacher become a mere servant? This could be claimed in the case of some TGs, which may "take the teacher step by step through every stage of every unit"

(Cunningsworth, 1984:52). However, the reality is that many TGs remain "no more than student editions with an inserted answer key" (Sheldon, 1987:3), and in this case the TG can be viewed as merely a weak servant of the textbook.

Compromises between these two extremes do exist of course, with some TGs providing detailed plans for sample lessons or units only, from which teachers can develop their own plans for later lessons or units.

Coleman (1985:84) argues in favour of the 'strong' TG "which can bolster the (NNS) teacher's confidence" and we would do well to maintain a global view and an awareness that the functions of the TG for NNS teachers "in the developing world will differ ... from those ... for teachers who have ample training, rich resources and small classes." (Cunningsworth and Kusel, 1991:129). The view of Richards, that teachers are 'deskilled' when their "decisions are largely based on the textbook and the teachers' manual" (1993:7) appears less damning when seen in the light of this awareness - I would argue that a given TG may be capable of 'deskilling' some (trained, western) teachers, while also being capable of 'skilling' others (i.e. 'untrained', NNS teachers).

The existing literature

Literature concerning evaluation of TGs is extremely thin on the ground (Coleman, 1985:85; Cunningsworth and Kusel, 1991:128). Materials reviews give scant attention to TGs, or ignore them completely (Coleman, 1985:86). This paper will necessarily also draw on the literature of general ELT materials evaluation.

Evaluation process and criteria

Why evaluate TGs?

Materials are not simply the everyday tools of the language teacher, they are an embodiment of the aims, values and methods of a particular teaching/learning situation. As such the selection of materials probably represents the single most important decision that the language teacher has to make. (Hutchinson, 1987:37)

This must especially be the case with TGs since they embody, (or should embody) aims, values and methods in an explicit manner. If this does not answer the question "Why evaluate TGs?" then Coleman's observation that "frequently ... TGs do not satisfy their (the teachers') needs, ... the guides are not guiding them" (1985:94) ought to do so.

The process of evaluation

We may identify three basic kinds of ELT materials evaluation; an intuitive, impressionistic approach, a formal prior-to-use evaluation, and a 'process' approach. The informal, impressionistic approach may be based on 'first impressions', perhaps gathered simply by flicking through materials or by reading publicity blurb. Even if a more thorough examination is carried out, it will not be comprehensive or in any way systematic. The limitations of this approach will be clear to anyone who has carried out such an evaluation only to be disappointed by later discoveries or results. These limitations are more fully discussed by Ellis (1997:37) and Low (1987:19).

The second approach is widely advocated, for example by Chambers (1997:29, 31). It is considered systematic, detailed, principled and comprehensive, involving, as it does, the use of carefully developed procedures and checklists of criteria which are used to perform a step-by-step examination of the materials. A system whereby points or marks are awarded may be utilised, which has the advantage of ensuring a degree of objectivity.

For the third approach, Ellis (1997:36) outlines a procedure involving predictive evaluation, choice of materials, followed by a post-use retrospective evaluation which may lead to further predictive evaluations. These stages equate with the 'input' and 'throughput' stages of Sheldon's (1987:5) tripartite schema, the third stage of which, 'output', refers to evaluation according to eventual learning outcomes. This procedure is supported by the general thrust of the argument in Rea-Dickins and Germaine (1993:145-152). This approach would seem more principled than the prior-to-use evaluation alone and as such has several advantages. Firstly, the implicit stress on needs analysis tends to ensure validity (Pilbeam, 1987:120). Secondly, a larger number of concerned individuals are likely to be involved in the evaluation. Finally, in-

class retrospective evaluation, which could involve empirical data, will tend to confirm or refute the validity of the initial stages of evaluation.

Checklists of criteria

Any systematic evaluation will involve the use of formalised lists of criteria. The important thing to remember here is that the quality of the questions asked is more crucial than the sheer number, as Dougill (1987:32) notes.

Unfortunately, there exist, to the writer's knowledge, only two published checklists for TG evaluation, those of Coleman (1985) and Cunningsworth and Kusel (1991). Appendix one summarises these in tabular form.

By contrast, checklists for evaluation of ELT materials in general exist in plenty. Eleven of these are summarised in tabular form in appendix two. These are included because, as we will see later, certain criteria may also be relevant for TG evaluation.

Teachers' guide evaluation in a specific context

As mentioned earlier, asking the right questions in an evaluation is obviously central to its success or failure. The ability to ask the right questions depends to a large extent on how context-specific the evaluation design is (Sheldon, 1987:6; Sheldon, 1988: 242; Williams, 1983: 253).

The remainder of this article discusses evaluation design and criteria in a specific context, a private language school offering general EFL conversation classes. It should be stressed that such evaluations are always context-dependent, and while the design arrived at will not be directly applicable in other contexts, it will be a relatively simple matter to make the necessary adjustments to the checklists and overall design. I have drawn attention in the text to several points at which such changes may be appropriate.

Evaluation design

From the point of view of trainers, teachers and learners, we may surmise that the greater their involvement in the evaluation process, the more valid the results of the evaluation will be (Chambers, 1997:29; Williams, 1981:155). However, things may

not be quite so straightforward. In my experience learners will enjoy playing a role only if the boundaries of their impact are clearly defined. If they are given 'too much'

responsibility for evaluation, many learners see teachers, trainers and management as shirking their own responsibilities.

A scheme for the involvement of learners in evaluation modified from, for example, Breen and Candlin (1987:26,27) would seem a good place to begin. In the absence of the need to prepare learners for any formal exams, this could be seen as a useful form of needs analysis, the results of which would feed into the evaluation criteria used.

Stage two could be a formalised student needs analysis such as the one exemplified by Matthews (1985:203,204) performed by the school director. Since we are aiming to select a TG, and not simply course materials, the school director will also need at this point to identify the fundamental needs of the teachers who are to use the TG.

The next stage, would perhaps be an impressionistic evaluation of the possible TGs. It should be noted at this point that a range of broadly suitable courses will have been previously identified through use of a published evaluation checklist, or one developed specifically for the given context, along with the results of the needs analysis. The impressionistic TG evaluation would be carried out by practising teachers. (Matthews, 1985:204). As well as the benefits for management of input from 'grassroots level', teachers will feel involved in the evaluation. A positive spin-off is teacher development: this engagement with 'training materials' having a potential teacher-training effect. This teacher involvement would go some way towards establishing face validity of the entire evaluation process (Pilbeam, 1987:120). At this point, if consensus is clear, some TGs (and their associated courses) may be eliminated, only successful ones going forward to the next stage.

The use by teachers, trainers and centre directors of a formal list of evaluation criteria would come next. Ideally this would also leave room for intuition and not be unnecessarily complex (Chambers, 1997:30,31).

Finally, since no 'ideal' TG is likely to emerge, due to the involvement of so many factors and stakeholders in the evaluation, the 'best fits' will be selected for piloting along with their associated coursebooks and ancillaries. Ellis (1997:40) reminds us to avoid conflating evaluation conclusions with evaluation recommendations and we should keep an open mind (and open eyes and ears) during piloting, in addition to using questionnaire and interview or diary techniques with the teachers involved to complete our TG evaluation. Student input concerning the associated textbook etc. may be sought through questionnaires.

This extended process should allow all stakeholders to feel involved to an appropriate degree and a reliable selection of TG (and associated materials) to be made.

Teachers' guide evaluation criteria

The following list of TG evaluation criteria, developed for conversation classes at a specific language school, draws partly on Cunningsworth and Kusel (1991), with additions from Coleman (1985) (both summarised in appendix one), and limited input from a several of the textbook and materials evaluation checklists outlined in appendix two. Two lists of questions are given, for 'global' and 'detailed' stages of evaluation. Global questions are more fundamental, and should be applied first. Items marked with an asterisk may be omitted or amended if they are considered to be incompatible with the context under scrutiny. It should also be noted here that an evaluation procedure for other kinds of courses, for example examination courses, or courses designed to develop single skills such as listening, would necessarily include additional questions appropriate to the course in question, and exclude others.

Global evaluation:

- Is the TG and its ancillaries readily available at an acceptable price, both to the school and to the learners?
- Does the TG provide a viable rationale for the information and guidance it provides?
- Does the TG make unreasonable assumptions about the users' knowledge and experience of language teaching?
- Does the TG deepen users' understanding of language teaching principles as they develop their practical teaching skills?

- Does the TG focus on one variety of English to the exclusion of others?*
- Is the advice given on teaching procedures explicit enough?
- Is there enough cultural information to enable teachers to interpret appropriately the situations represented in the teaching materials?*
- Is the guidance provided unreasonably prescriptive?
- Does the TG have helpful things to say about the language learning process which are also supported by modern theory and research?
- Does the TG make explicit an awareness of different learning styles and strategies and suggest ways of using and developing them?
- Does the TG consider a variety of roles for the teacher with reference to the nature of each learning activity?

Detailed evaluation:

- Does the TG explicitly inform the teacher how it can best be used?
- Are the objectives of the course set out clearly and rationally?
- Is the basic teacher-student relationship implied in the TG, and the content and methods contained within it appropriate to the context of use?
- Is the TG likely to help heighten and sustain learner motivation?
- Is use of the TG likely to result in lessons the learners will enjoy? (Ellis and Ellis, 1987:96).
- Is the TG free of the use of confusing metalanguage?
- Do the design and content of the TG suggest ease of use? (Ellis and Ellis, 1987:91)
- Does the TG aim to maximise learners' opportunities to develop sociopragmatic and strategic competencies? (Cunningsworth, 1984:47-51).
- Does the TG suggest appropriate ways for the teacher to evaluate each activity, lesson, and sequence of lessons?
- Is the information about, and guidance on handling language items adequate, unambiguous and appropriate? Are there any helpful notes about potential problems which may arise regarding language items in this context?
- Does the TG predict difficulties in understanding the cultural setting and background in the materials, and provide sufficient information about, and explanation of them?*

- Does the TG provide clear but adequately flexible guidance in selecting and sequencing units, planning them into a scheme of work and integrating them into the programme as a whole?
- Does the TG suggest procedures for the planning, preparation and conduct of lessons in an appropriate manner and in sufficient detail?
- Does the TG suggest alternative routes through activities, lessons or units which can be helpful when things do not go according to plan?
- Does the TG advise the teacher when and how to correct students' language and about the likely responses of learners in this context to various kinds of correction?
- Does the TG provide clear and unambiguous answers to tasks set?
- Does the TG provide adequate guidance for the checking of learning, both formally and informally?

Final evaluation design for a private language school context

The following outline for TG (and by implication materials in general) evaluation is what I see as a pragmatic compromise based on my own very personal experience of what is actually possible in practice in the private language school context. Again, it should be borne in mind that this particular design is conditioned by the context and that it will ideally to be customised according to need.

1. Marking boundaries:

Design of, and criteria for evaluation are mutually agreed, (and if necessary put in writing, and signed) by owners, upper management, centre directors (directors of studies) and managers (business managers).

2. Impressionistic evaluation:

Practising teachers of all levels of experience within the school (or schools if the context is a chain or network) rank TGs under consideration from best to worst (norm referenced).

3. Formal evaluation:

Trainers, centre directors and selected teachers evaluate the TGs as: good = 3 points, acceptable = 1 point, and fail = 0 points against each of the

criteria (criterion referenced). Scores are collated and the cut off point decided.

4. Piloting:

The TGs still under consideration are piloted along with their coursebooks and ancillaries at several branches by teachers who can be relied upon to 'teach from the book'. The teachers are paid to write evaluation reports mid-course and at the end of the course.

5. Final selection:

By consensus among the parties in 1. above, bearing in mind the results of 2., 3. and 4.

If a suitable TG, textbook, workbook and tape set can be found through a systematic evaluation procedure of this kind, the benefits are likely to be several. Firstly, training budgets may be reduced somewhat, with trainers being left free to concentrate on fine tuning instead of battling with basic problems caused by unsuitable TG's and materials. In addition, teachers are more likely to feel more comfortable with the tools of their trade due to their active involvement in the process of evaluation and selection. Teachers may thus feel more value in adapting course materials and developing new ones to cover any 'weak spots' in the course and TG selected.

References

Breen, M.P. and C.N. Candlin (1987). 'Which Materials?: A Consumers' and Designers' Guide.' in Sheldon, L. E. (ed.).

Coleman, H.(1985.) 'Evaluating teachers' guides: Do teachers' guides teach teachers?' in Alderson, J.C. (ed.). *Evaluation. Lancaster Practical Papers in English Language Education. Vol. 6.* Oxford: Pergamon.

Chambers, F. (1997). Seeking consensus in coursebook evaluation.' *ELT Journal.* 51/1.

Cunningsworth, A. (1984). *Evaluating and selecting EFL teaching materials.* London: Heinemann.

Cunningsworth, A. and P. Kusel (1991). 'Evaluating teachers' guides.' *ELT Journal.* 45/2.

Dougill, J. (1987). 'Not so obvious.' in Sheldon, L. E. (ed.).

- Ellis, R.** (1997). 'The empirical evaluation of language teaching materials.' *ELT Journal*. 51/1.
- Ellis, J. and P. Ellis.** (1987). 'Learning by design: Some criteria in EFL coursebooks.' in Sheldon, L. E. (ed.).
- Grant, N.** (1987). *Making the most of your textbook*. London: Longman.
- Harmer, J.** (1991). *The Practice of English Language Teaching*. London: Longman.
- Hutchinson, T.** (1987). 'What's underneath?: An interactive view of materials evaluation.' in Sheldon, L. E. (ed.).
- Littlejohn, A. and S. Windeatt** (1988). Beyond language learning: perspectives on materials design. In R. K. Johnson (Ed.), *The Second Language Curriculum*. Cambridge: Cambridge University Press.
- Low, G.** (1987). 'The need for a multi-perspective approach to the evaluation of foreign language teaching materials.' *Evaluation and Research in Education*. 1/1.
- Matthews, A.** (1985). 'Choosing the best available textbook.' in Matthews, A. et al. *At the Chalkface*. London: Arnold.
- McDonough, J. and C. Shaw** (1993). *Materials and Methods in ELT, a teacher's guide*. Oxford: Blackwell.
- Pilbeam, A.** (1987). 'Can published materials be widely used for ESP courses?' in Sheldon, L. E. (ed.).
- Rea-Dickins, P. and K. Germaine** (1992). *Evaluation*. Oxford: Oxford University Press.
- Richards, J.C.** (1993). 'Beyond the textbook: The role of commercial materials in language teaching.' *RELC Journal*. 24/1.
- Sheldon, L.E. (ed.)** (1987). *ELT Textbooks and Materials: Problems in Evaluation and Development. ELT Documents 126*. Oxford: Modern English Publications / The British Council.
- Sheldon, L.E.** (1988). 'Evaluating ELT textbooks and materials.' *ELT Journal*. 42/4.
- Williams, D.** (1983). 'Developing criteria for textbook evaluation.' *ELT Journal*. 37/3.
- Williams, R.** (1981). 'A Procedure for ESP Textbook Analysis and Evaluation on Teacher Education Courses.' *The ESP Journal*. 1/2.

Appendix 1 – A Summary of Two Instruments for the Evaluation of Teachers' Guides

	Coleman 1985	Cunningworth and Kusel 1991
Total number of evaluation factors considered (number of evaluation questions)	10	31
Number of categories of factors	5	12 (Global' evaluation factors = 3; 'Detailed' factors = 9)
Number of factors (questions) per category	<ul style="list-style-type: none"> • Primary factors = 2 • Factors concerning presentation of content = 2 • Implementation factors = 2 • Evaluation factors = 2 • Presentation factors = 2 	<p>Global evaluation factors concerning...</p> <ul style="list-style-type: none"> • General criteria = 8 • Language and language learning = 5 • development of teachers' awareness of theory = 2 <p>Detailed evaluation factors concerning...</p> <ul style="list-style-type: none"> • Objectives and content = 2 • Cultural loading = 2 • Procedural guidance = 2 • Advice about the unpredictable = 1 • Correction and testing = 4 • Motivation = 1 • Presentation and use = 2 • TGs not in English = 1 • Lesson evaluation = 1

Appendix 2 - A Summary of Eleven Checklists for the Evaluation of ELT Textbooks and Materials

Checklist of criteria involves	Breen and Candlin 1987	Cunningsworth 1984	Dougill 1987	Grant 1987	Harmer 1983	Littlejohn and Windeatt 1989	Matthews 1985	McDonough and Shaw 1993	Richards 1993	Sheldon 1988	Williams 1983
TB or materials or both?	Both	Both	Both	TB	Both	Both	TB	Both	TB	Both	TB
context: ESL / EFL / ESP?	All	ESL/EFL	ESL/EFL	ESL/EFL	ESL/EFL	ESL/EFL	ESL/EFL	All	ESL/EFL	ESL/EFL	ESL (EFL)
explicit impressionistic preliminary evaluation?	No	No	No	No	No	No	Yes	No	No	No	No
practising teachers explicitly?	Yes	Yes*	Yes*	Yes*	Yes*	No	No	No	Yes*	Yes*	No
learners?	Yes	No	Yes*	Yes*	No	No	No	No	No	No	No
other stakeholders?	Yes*	No	No	Yes*	No	No	No	No	No	No	No
number of evaluation criteria (questions)	34 + 13 (+16 Qs. for learners)	54	51	'CATALYST' framework = 14 Qu'aires = 30	22	6	19	31	29	17	27
explicit needs analysis framework?	Yes	Yes**	No	No	Yes	No	Yes	No	No	No	No

NB: since several of the above checklists were set forth as preliminary investigations and not as complete working instruments, the above table should not be seen as in any sense an evaluative statement.

* Includes criteria which can **only** be judged by practising teachers, learners or other stakeholders.

** Discussed in chapter 1, although framework is not actually developed.