Evaluating ELT textbooks and materials

Leslie E. Sheldon

ELT coursebook publishing is a multi-million pound industry, yet the whole business of product assessment is haphazard and under-researched. Coursebooks are often seen by potential consumers—teachers, learners and educational purchasers—as market ephemera requiring invidious compromises between commercial and pedagogical demands. Some practical and theoretical reasons for such grassroots discontent are discussed, as are previous textbook studies and qualitative 'checklists'. The state of the informational assistance available to intending purchasers is also examined. Finally, a set of 'common core' qualitative criteria is advanced, whose purpose it would be to make evaluation and selection more systematic and informed.[1]

According to Hutchinson and Waters, textbook evaluation is basically a straightforward, analytical 'matching process: matching needs to available solutions' (Hutchinson and Waters 1987:97). My own view is that this issue is rather more emotive and controversial for teachers; many would agree with Swales (1980) that textbooks, especially coursebooks, represent a 'problem', and in extreme cases are examples of educational failure. I would like to explore the reasons for such strong reactions, and to put forward possible evaluative solutions. I wish to concentrate on coursebooks because, whether we like it or not, these represent for both students and teachers the visible heart of any ELT programme. The selection of a particular core volume signals an executive educational decision in which there is considerable professional, financial and even political investment. This high profile means that the definition and application of systematic criteria for assessing coursebooks are vital. Supplementary textbooks and materials on the other hand may not carry the same burden. The evaluative criteria for these can to some extent remain implicit, or be allowed to define themselves more informally in the local situation.

Classroom attitudes
ELT coursebooks evoke a range of responses, but are frequently seen by teachers as necessary evils. Feelings fluctuate between the perception that they are valid, labour-saving tools, and the doleful belief that 'masses of rubbish is skilfully marketed' (Brumfit 1980:30). In basic terms, there seems to be a 'coursebook credibility gap' (Greenall 1984:14) because of emphatic contradictions and potential conflicts of interest in their creation, commercial exploitation, public assessment, selection, and ultimate classroom use. ELT books are frequently seen as poor compromises between what is educationally desirable on the one hand and financially viable on the other. In simple terms, they often do not seem to provide good value for money.

In the attempt to make the volumes in question seem suitable for all learners in all situations, excessive claims by some authors and publishers have made such disaffection inevitable. Staffroom conversations about which parts of a specific textbook can be salvaged for actual teaching purposes are a constant reminder of the belief, publisher's blurbs notwithstanding, that 'The whole business of the management of language learning is far too complex to be satisfactorily catered for by a pre-packaged set of decisions embodied in teaching materials' (Allwright 1981:9).

The sheer labour-intensiveness of developing classroom materials, the pressures of heavy timetables, and the highly restrictive nature of most teaching situations nevertheless force the teacher (or educational purchaser) to rein in his or her reservations, and to choose a book which only approximates to the needs of the local context. It is not always a happy compromise, and is most often made in haste and with a paucity of systematically applied criteria. Coursebooks are perceived by many to be the route map of any ELT programme, laying bare its shape, structure, and destination, with progress, programme, and even teacher quality being assessed by learners in terms of sequential, unit-by-unit coverage. In this situation, wrong selection can be a particularly keen reason for regret. When a textbook is imposed on both parties by a higher authority, and when there is no possibility of change or modification, the discontent is no less acute for being futile.

It is a cruel paradox that for students, teacher-generated material (which potentially has a dynamic and maximal relevance to local needs) often has less credibility than a published textbook, no matter how inadequate that may be. Against the public endorsement implied by printed covers, home-grown materials appear in a poor light. This is especially true for the seemingly disconnected snippets of authentic texts or teacher-created worksheets that are a consequence of the 'communicative' approach. Quite simply, it would seem that coursebooks exert a kind of 'backwash' effect. This is particularly true where their form and content is restricted because programmes are linked to an English Language qualification, probably the most common ELT situation worldwide.

These frustrations can be assuaged, at least potentially, when local textbooks are developed in order to accommodate the on-the-spot needs of learners and preferences of teachers. For the future, one might speculate that with the expansion in desk-top publishing via microcomputers, even greater scope for interaction could one day be possible, with the textbook as we know it eventually disappearing; it might be replaced instead by a 'published' core materials program, which the teacher could modify and supplement as required, hard copies being placed between 'glossy' textbook covers to satisfy student expectations about format and about quality of production.[2]

But the fact is that coursebooks are *here*, and are even exploited in traditional environments as a method of controlling large classes of learners. Such pedagogical experiences generate expectations about what a coursebook should contain, what it should look like, and how it should be used. These perceptions are frequently carried over into environments where more freedom and choice are in reality possible.

Common theoretical and practical problems

Other, more immediate reasons for disappointment seem partly to do with a variety of common design flaws on one level, and a scepticism about the theoretical premises of many coursebooks on another: all this despite the

Leslie Sheldon

fact that 'to put a book on the open market implies a moral contract that the book has been cleared of basic faults' (Brumfit 1980:30). That the same complaints come up repeatedly suggests a violation of this implicit agreement, and a substantive mismatch between what is produced and what teachers would like to have. Of course, the situation is not helped by the fact that textbooks are necessarily static and a little outdated, not least because of the long delays between writing and publication. The feedback loop for homegrown materials is of course much shorter and quicker.

Practical problems include such recurring faults as the failure to describe adequately the language levels of target learners (the indiscriminate use of terms like 'beginner' and 'intermediate' abounds); the production of workbooks or exercise pages that demand microscopic handwriting; the omission of course rationales stating exactly who the book is intended for, or how the material is selected and sequenced; a surrender to the economic pressures that demand a maximum textual density on each page; and the production of teacher's books that are little more than student editions with inserted answer keys. The purchaser is also confronted with a terminological looseness (for example, 'communicative', 'authentic', 'notional/functional', 'lexical syllabus', etc.) that makes meaningful comparisons of textbooks difficult.

Publishers sometimes neglect matters of cultural appropriacy; they fail to recognize the likely restrictions operative in most teaching situations; and they are not always aware of the pedagogical implications of current theory and research in linguistics and language learning. Many textbooks, for example, use artificial, whole-sentence dialogues, despite the descriptions available of the truncated nature of authentic oral interaction (Cunningsworth 1987). A reader's use of schemata (Carrell 1984) and text sampling is often ignored in 'Reading Skills' books which encourage linear, word-by-word, bottom-up processing via visually intrusive comprehension questioning (often in the margins) and arrow-drawing 'reference' exercises. The selection and presentation of vocabulary too often seem to be accomplished without system in some coursebooks, despite the relevant studies and sources of information to which reference could be made (for example, lexical field theory, componential analysis, West's *A General Service List of English Words*, and the *Cambridge English Lexicon*).

There may indeed be a closed circle at work here, wherein textbooks merely grow from and imitate other textbooks and do not admit the winds of change from research, methodological experimentation, or classroom feedback.

Thus, for good procedural and theoretical reasons, textbooks are frequently seen as the tainted end-product of an author's or a publisher's desire for a quick profit. Disjointed material produced in specific and possibly limited situations is generalized and stitched together under flashy covers; or so it often seems. The large number of single-edition, now defunct coursebooks produced during the past ten years testifies to the market consequences of teachers' verdicts on such practices. In the United States alone, 28 major publishers now offer 1,623 ESL textbooks between them (Goodman and Takahashi 1987). Purchasers are rightly sceptical about the improved utility or methodological 'advance' claimed for each new publication.

Previous evaluative proposals

Teachers are consumers, just like students or educational administrators. All these groups, of course, can have potentially conflicting notions of what

a good textbook should be. But whatever they feel, where can they all turn for advice on how to make informed decisions about the best value for money?

The literature on the subject of textbook evaluation is not very extensive. Various writers have suggested ways of helping teachers in particular to be more sophisticated in their evaluative approach, by positing 'checklists' based on supposedly generalizable criteria. These sometimes elaborate systems use a variety of 'scoring' methods to assess how well specific textbooks measure up under scrutiny (see Tucker 1975, van Lier 1979, Allwright 1981, and Williams 1983). Tucker, for example, proposes an ingenious method whereby textbooks are assigned numerical scores, which are then plotted on a 'Value Merit Product Graph', the object of which is to compare the resulting score curve against an ideal target profile drawn up in advance by the teacher.

Of course, the criteria and key questions central to such schemes partly depend on the swings of linguistic fashion. Nowadays, one would probably not rate 'adequacy of pattern inventory' (Tucker 1975:360) as highly as a few years ago. Such decisions would, however, depend on one's own local priorities and preferences. Over the years, the relative importance of different criteria would change, along with the interpretation given to the scores assigned in each category. A solid structurally-based coursebook might not pass muster quite as well as it once did. Moreover, as with all the attempts to objectify what are essentially subjective phenomena, some of the categories would in any case be exceedingly difficult to quantify. Factors such as 'Competence of the author' (Tucker 1975:360) and whether or not a book is 'based on a contrastive analysis of English and L1 sound systems' (Williams 1983:255) might present problems of clarification and scoring.

Though providing food for thought, checklists and questionnaires like these have probably had little real influence on textbook evaluation in terms of either ELT reviewing practice or educational decision making. Perhaps they have simply not had the currency they deserve, most teachers, at any rate, being unaware that they even exist. Finding back issues of relevant journals in many teaching contexts (and indeed in some ELT libraries) is also very difficult. The subject bibliography is short, so there are few published discussions for the interested teacher to consult, even if he/she had time to do so. Perhaps this explains why coursebook evaluation is usually so *ad hoc* and rushed, with teachers trying to make choices on the basis of such simplistic criteria as 'popularity', in the belief that if a book sells well, somehow, somewhere, someone must be doing something right.

Obviously, the articles and checklists that exist do not reflect the whole story; all TEFL/TESP teacher training programmes generate their own approaches to coursebook evaluation (e.g. Cunningsworth 1979), but because of the isolationist nature of TEFL, these approaches vary greatly in their rigour, consistency, utility, and awareness of relevant scholarship. There would seem to be a point at which celebration of 'welcome diversity' (Bloor, Swales and Williams 1984:4), especially in tertiary-level teacher training becomes a smug acceptance of entrenched confusion, rather than a positive basis on which to initiate methodological development and coherent change. Nowhere is this state of affairs clearer than in the uneven quality of the evaluative tools given to the thousands of EFL/ESP teachers trained in the UK each year.

Leslie Sheldon

The reviews spread across the ELT press also seem unlikely to influence actual coursebook selection. Their discursive format makes it difficult to separate description, guidance, and criticism. Also, they seem for the most part to be unruffled by schemes such as those put forward by Tucker and others, which aim for a graphically presented, concise mode of textbook analysis. Trying to ascertain, for example, how many units a coursebook has, or the number of hours allocated to each unit, often involves considerable skimming. It is difficult, in a number of cases, to say with certainty exactly what a reviewer really does feel about the relative advantages/disadvantages of a specific coursebook.

Reviews, like the books they dissect, vary in their academic and professional quality, and in their ability to be of practical assistance. The approach often tends to be idiosyncratic (perhaps no bad thing), and reviewers can often be constrained by the desire either to be polite, or to be entertaining at the expense of fairness. This is all made worse by the fact that reviewers are a transient lot.

Even assuming that the teacher or educational purchaser is able to peruse the ELT press, he or she can obtain only a sporadic impression of what is currently available. In the event that the consumer finds a potentially interesting title in a publisher's catalogue or on the shelves of a bookshop during a visit to the UK, it is difficult to discover when, where, or if it has been previously assessed. Old reviews pass quickly from the scene, except where publishers revive positive ones for marketing purposes.

As Christopher Brumfit observes: 'There is no *Which* for textbooks' (Brumfit 1980:30).[5] Perhaps it is high time there was. Though the *ELT Update* (produced by KELTIC) is a mine of potentially useful, carefully structured information, its circulation is small and its publication frequently delayed. Reviews in the large-circulation periodicals do not present the grid-style tables typical of *Which?*, preferring to use the traditional conflation of descriptive information and subjective commentary. The notable exception to this is *ELT Journal*, which uses an evaluative table at the end of survey reviews. This system has another advantage in that it allows similar textbooks to be compared, rather than to be scrutinized only as individual, isolated objects of analysis, which is typical of most reviews.

The physical appearance of textbooks and perhaps even the extent to which they provide guidance for their myriad users have changed substantially over the past few years, if only partially in response to teachers' demands. We might now expect a corresponding change in reviews as well, so that they can fulfil a genuine advisory role. The recent experimentation in the *EFL Gazette* (see Issue 96 p.7; and Issue 97, p.15) with 'student-generated' and 'teacher-consensus' reviews would seem to be steps in the right direction. At the very least, we might hope for the production of a compendium of reviews culled from various sources so that teachers can have access to what has been thought and said about various books over a period of time.

For discussion, I would like to present what I think is a 'bell-jar' summary of common-core factors that reviewers, administrators, teachers, learners, and educational advisers most frequently use in deciding whether or not a textbook is chosen. A major caveat is necessary, however: no one is really certain what criteria and constraints are actually operative in ELT contexts worldwide, and textbook criteria are emphatically local. Not all the criteria described would be deployed simultaneously, nor is the list definitive.

Consumers would obviously emphasize other factors that relate specifically, and perhaps dramatically, to their own unique situations.

The point is, of course, that any culturally restricted, global list of criteria can never really apply in most local environments, without considerable modification. We can be committed only to checklists or scoring systems that we have had a hand in developing, and which have evolved from specific selection priorities. So we need some points around which our thoughts can crystallize.

The elements presented might perhaps lead to revamped formats for ELT book reviews, or, on a more modest level of optimism, to the development of evaluation sheets (see Figure 1) which could be used in ELT departments to record in-house coursebook assessments. Ideally this evaluative framework could also have a bearing on the appraisal of teacher-produced materials, because identifying a qualitative yardstick for published coursebooks, regardless of whether or not the teacher has the free-

FACTUAL DETAILS

Title: ....
Author(s): ....
Publisher: .... Price: ....
ISBN: .... No. of Pages: ....
Components: SB/TB/WB/Tests/Cassettes/Video/CALL/Other ....
Level: .... Physical size: ....
Length: .... Units .... Lessons/sections .... Hours ....
Target skills: ....
Target learners: ....
Target teachers: ....

ASSESSMENT (* Poor   ** Fair   *** Good   **** Excellent)

Factor	Rating and comments
Rationale	
Availability	
User definition	
Layout/graphics	
Accessibility	
Linkage	
Selection/grading	
Physical characteristics	
Appropriacy	
Authenticity	
Sufficiency	
Cultural bias	
Educational validity	
Stimulus/practice/revision	
Flexibility	
Guidance	
Overall value for money	

Figure 1: Textbook evaluation sheet

Leslie Sheldon

dom to choose what he or she uses in class, would seem to be central to coping with 'gaps' via subsequent supplementation and adaptation.

There are several key questions to ask ourselves about each feature:

Rationale —Why was the book written in the first place, and what gaps is it intended to fill?
—Are you given information about the Needs Analysis or classroom piloting that were undertaken?
—Are the objectives spelt out?

Availability —Is it easy to obtain sample copies and support material for inspection?
—Can you contact the publisher's representatives in case you want further information about the content, approach, or pedagogical detail of the book?

User definition —Is there a clear specification of the target age range, culture, assumed background, probable learning preferences, and educational expectations?
—Are entry/exit language levels precisely defined, e.g. by reference to international 'standards' such as the ELTS, ACTFL or Council of Europe scales, or by reference to local or country-specific examination requirements?
—In the case of an ESP textbook, what degree of specialist knowledge is assumed (of both learners and teacher)?

Layout/graphics —Is there an optimum density and mix of text and graphical material on each page, or is the impression one of clutter?
—Are the artwork and typefaces functional? colourful? appealing?

Accessibility —Is the material clearly organized?
—Can the student find his or her location in the material at any point, i.e. is it possible to have a clear view of the 'progress' made, and how much still needs to be covered?
—Are there indexes, vocabulary lists, section headings, and other methods of signposting the content that allow the student to use the material easily, especially for revision or self-study purposes?
—Is the learner (as opposed to the teacher) given clear advice about how the book and its contents could be most effectively exploited?

Linkage —Do the units and exercises connect in terms of theme, situation, topic, pattern of skill development, or grammatical/lexical 'progression'?
—Is the nature of such connection made obvious, for example by placing input texts and supporting exercises in close proximity?
—Does the textbook cohere both internally and externally (e.g. with other books in a series)?

Selection/grading —Does the introduction, practice, and recycling of new linguistic items seem to be shallow/steep enough for your students?
—Is there a discernible system at work in the selection and grading of these items (e.g. on the basis of frequency counts, or on the basis of useful comparisons between the learner's mother tongue and English)?
—Is the linguistic inventory presented appropriate for *your* purposes, bearing in mind the L1 background(s) of your learners?

Evaluating ELT textbooks   243

Physical characteristics
—Is there space to write in the book?
—Is the book robust? too large? too heavy?
—Is the spine labelled?
—Is it a book that could be used more than once, especially if it is marked by previous students?

Appropriacy
—Is the material substantial enough or interesting enough to hold the attention of learners?
—Is it pitched at the right level of maturity and language, and (particularly in the case of ESP situations), at the right conceptual level?
—Is it topical?

Authenticity
—Is the content obviously realistic, being taken from L1 material not initially intended for ELT purposes?
—Do the tasks exploit language in a communicative or 'real-world' way?
—If not, are the texts unacceptably simplified or artificial (for instance, in the use of whole-sentence dialogues)?

Sufficiency
—Is the book complete enough to stand on its own, or must the teacher produce a lot of ancillary bridging material to make it workable?
—Can you teach the course using only the student's book, or must all the attendant aids (e.g. cassettes) be deployed?

Cultural bias
—Are different and appropriate religious and social environments catered for, both in terms of the topics/situations presented and of those left out?
—Are students' expectations in regard to content, methodology, and format successfully accommodated?
—If not, would the book be able to wean students away from their preconceived notions?
—Is the author's sense of humour or philosophy obvious or appropriate?
—Does the coursebook enshrine stereotyped, inaccurate, condescending or offensive images of gender, race, social class, or nationality?
—Are accurate or 'sanitized' views of the USA or Britain presented; are uncomfortable social realities (e.g. unemployment, poverty, family breakdowns, racism) left out?

Educational validity
—Does the textbook take account of, and seem to be in tune with, broader educational concerns (e.g. the nature and role of learning skills, concept development in younger learners, the function of 'knowledge of the world', the exploitation of sensitive issues, the value of metaphor as a powerful cognitive learning device)?

Stimulus/practice/ revision
—Is the course material interactive, and are there sufficient opportunities for the learner to use his or her English so that effective consolidation takes place?
—Is the material likely to be retained/remembered by learners?
—Is allowance made for revision, testing, and on-going evaluation/marking of exercises and activities, especially in large-group situations; are ready-made achievement tests provided for the coursebook, or is test development left for the hardpressed teacher? Are 'self-checks' provided?

Flexibility
—Can the book accommodate the practical constraints with which you must deal, or are assumptions made about such things as the availability

244   *Leslie Sheldon*

of audio-visual equipment, pictorial material, class size, and classroom geography; does the material make too many demands on teachers' preparation time and students' homework time?

—Can the material be exploited or modified as required by local circumstances, or is it too rigid in format, structure, and approach?

—Is there a full range of supplementary aids available?

Guidance —Are the teacher's notes useful and explicit?

—Has there been an inordinate delay between the publication of the student's and teacher's books which has meant that teachers have had to fend for themselves in exploiting the material?

—Is there advice about how to supplement the coursebook, or to present the lessons in different ways?

—Is there enough/too much 'hand-holding'?

—Are tapescripts, answer keys, 'technical notes' (in the case of ESP textbooks), vocabulary lists, structural/functional inventories, and lesson summaries provided in the Teacher's Book?

—Is allowance made for the perspectives, expectations, and preferences of non-native teachers of English?

Overall value for —Quite simply, is the coursebook cost-effective, easy to use, and successful
money in your teaching situation, in terms of time, labour, and money?

—To what extent has it realized its stated objectives?

However the above categories might be 'scored' (with pluses and minuses, or stars, etc.), it is clear that different kinds of user could have divergent (and even opposing) concerns. For example, an educational administrator, with an eye to financial issues, might consider ratings on 'Sufficiency' or 'Physical characteristics' to be more important than those for 'Rationale'. Moreover, the same coursebook, when judged by the same criteria, could be 'successful' in one context, but not in another. For instance, the presence of cultural elements could be welcomed in a situation where the teacher felt that they were a central vehicle for learning English; on the other hand, they could be seen as irrelevant and even offensive in other teaching situations. Similarly, the 'hand-holding' provided by some teacher's books could be interpreted either as useful guidance or unwanted regimentation. Also, of course, ratings would clearly be influenced by whether or not instruction was carried on in ESL or EFL environments, and by such factors as the divergent claims of EFL versus ESP, wherein the textbook/homegrown materials mix can be radically different.

The nature of It is clear that coursebook assessment is fundamentally a subjective, rule-
coursebook of-thumb activity, and that no neat formula, grid, or system will ever
assessment provide a definitive yardstick. But at the very least, perhaps the use of similar evaluative parameters will help to make it, when time and circumstances allow, a more coherent, thoughtful enterprise than it often is at present.

Textbook appraisal is not a once-only activity. When a coursebook is selected, its success or failure can only be meaningfully determined during and after its period of classroom use. Learners are not taught in a vacuum, but come from somewhere and are proceeding towards specific educational goals and future training. The coursebook ultimately needs to be appraised

in terms of its integration with, and contribution to, these longer-term goals.

Finally, perhaps when more has been written and said about helping consumers to choose appropriate textbooks, when research discovers what criteria purchasers actually do use, these books will in turn become the kind of educational tools that consumers want. This will also be achieved by articulate expectations filtering back from the classroom, especially when teachers are afforded more effective feedback channels than they have at present. For the moment, we need to discover whether or not a *de facto* evaluative consensus exists at all, and whether there is any foundation upon which universal criteria could be erected. □

Received March 1988

Notes

1 This article is based on a paper delivered at TESOL (Miami), April 1987, and on a series of coursebook assessment workshops held in Germany and France, and at the universities of Aston, East Anglia, Warwick, Birmingham, and Surrey.

2 I am indebted to Michael Hoey for the concept of an 'electronic textbook'.

3 *Which?* is a consumer magazine published in the UK. Different products or services are compared on a grid by independent assessors, using a four-dot rating system.

References

Allwright, R. 1981. 'What do we want teaching materials for?' *ELT Journal* 36/1:5–18.

Bloor, M., J. Swales and D. Williams. 1984. 'ESP: strength in diversity.' *EFL Gazette* 56/4.

Brumfit, C. J. 1980. 'Seven last slogans.' *Modern English Teacher* 7/1:30–31.

Carrell, P. L. 1984. 'Schema theory and ESL reading: classroom implications and applications.' *Modern Language Journal* 68/4:332–43.

Cunningsworth, A. 1979. 'Evaluating course materials' in S. Holden (ed.): *Teacher Training*. London: Modern English Publications.

Cunningsworth, A. 1987. 'Coursebooks and conversational skills' in L. E. Sheldon (ed.): *ELT Textbooks and Materials: Problems in Evaluation and Development (ELT Documents 126)*. Oxford: Modern English Publications/The British Council.

EFL Gazette. 1987. 'Business books survey.' Number 96, page 7.

EFL Gazette. 1987. 'Five great reads?' Number 97, page 15.

Goodman, P. and S. Takahashi. 1987. 'The ESL textbooks explosion: a publisher profile.' *TESOL Newsletter* 4:49–51.

Greenall, S. 1984. 'The coursebook credibility gap.' *EFL Gazette* 53/54:14.

Hutchinson, T. and A. Waters. 1987. *English For Specific Purposes: A Learning-Centred Approach*. Cambridge: Cambridge University Press.

van Lier, L. 1979. 'Choosing a new EFL course.' *Mextesol Journal* III/3:2–14.

Swales, J. 1980. 'ESP: the textbook problem.' *ESP Journal* 1/1:11–23.

Tucker, C. A. 1975. 'Evaluating beginning coursebooks.' *English Teaching Forum* XIII/3/4:355–61.

West, M. 1953. *A General Service List of English Words*. London: Longman.

Williams, D. 1983. 'Developing criteria for textbook evaluation.' *ELT Journal* 37/3:251–5.

The author

Leslie Sheldon is Director of ELT, Pitman Education and Training Ltd. He has taught at schools, universities, and teacher-training establishments in Canada, Iran, Algeria, Italy, Sweden and the UK, as well as having been an ESP consultant. In 1989 Dr Sheldon will direct a British Council specialist seminar to be held in the UK, entitled 'ELT Textbooks And Materials: Evaluation, Exploitation, Adaptation and Design'.

246   *Leslie Sheldon*