A step forward: investigating expertise in materials evaluation

Keith Johnson, Mija Kim, Liu Ya-Fang, Andrea Nava, Dawn Perkins, Anne Margaret Smith, Oscar Soler-Canela, and Wang Lu

This article reports on a study investigating the textbook evaluation techniques of novice and experienced teachers, which was conducted by the Language Teaching Expertise Research Group (or LATEX) within Lancaster University’s Department of Linguistics and English Language. Three ELT teachers were chosen to evaluate the student and teacher editions of a newly-released ELT textbook using the technique of concurrent verbalization. The results of the research add to the growing body of knowledge on expertise, providing insight into the differences between the teachers with respect to their various evaluation strategies. They also point to implications for the development of teacher education and training.

Why this study?

In the last several years the number of English-language teaching materials on the market has grown exponentially, addressing a variety of learner interests, skill levels, and tastes. Among other features, these materials also vary in their linguistic design, focus, and objectives, making the choice of a textbook—an integral part of many ELT classrooms—a seemingly formidable task. Teachers or those training to become English teachers can look to a number of sources to help them to make more informed decisions when evaluation is required. For instance, some teacher training courses offer instruction in materials evaluation. In addition, there are various references in the literature that address the topic—Tomlinson (2003) and McGrath (2002) are two of the more recent books that include chapters on evaluation—but empirical studies revealing what experienced textbook evaluators actually do are rare.

In this small-scale case study of textbook evaluation conducted by LATEX at Lancaster University, the researchers chose to focus on the practices of teachers at different stages of their careers to discover what effect this variable might have on their evaluation styles. This was in the hope that the conclusions drawn from the research would help point to more effective means of evaluation, which can thereby be utilized by all teachers who are required to make informed decisions about choosing a coursebook. The researchers note that the scale of the study—with three participants—necessarily limits its generalizability but, nevertheless, it is hoped that
the research framework may also serve as a model for future large-scale research projects focusing on materials evaluation procedures.

**Methodology**

Why ‘think-aloud’ protocols?

This ‘collective case study’ of the textbook evaluation practices of experienced and novice teachers was designed to facilitate greater understanding of effective and efficient techniques of evaluation with the aim of ‘theorizing’ about these techniques in general (Stake 2000: 437). When formulating the design of the case study, the researchers decided that the technique that would best capture the introspective thought processes of the participants involved was concurrent verbalization. This is a technique that has been criticized for possibly distorting thinking processes, especially when the participants are asked to verbalize their thoughts retrospectively, or in a language other than their first language, or when they are asked to analyse their thoughts as they report them. None of these demands were placed on the informants in this study, nor were any time constraints imposed. Given the virtual impossibility of obtaining any truly complete records of complex thought processes, some of which will be hidden from the thinkers themselves, the use of ‘think-aloud’ protocols offers some insights that other methods would not be able to reveal. Even pauses in the verbalization can be revealing, indicating phases when intense thinking is taking place.

The participants

In total, three teachers took part in the study (referred to as T1, T2, and T3); their teaching experience ranged from one year (T1) to 12 years (T3) and their textbook evaluation training and/or experience ranged from none to that of the level of curriculum coordinator (see Table 1 for more details).

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Years teaching</strong></td>
<td>1 year</td>
<td>5 years</td>
<td>12 years</td>
</tr>
<tr>
<td><strong>Full- or part-time</strong></td>
<td>Part-time</td>
<td>Full- and part-time</td>
<td>Full-time</td>
</tr>
<tr>
<td><strong>ELT qualifications</strong></td>
<td>No</td>
<td>Cert TEFL</td>
<td>Cert TESOL, DELTA, MSc</td>
</tr>
<tr>
<td><strong>Previous textbook evaluation training</strong></td>
<td>No</td>
<td>No</td>
<td>Yes, as part of academic qualifications.</td>
</tr>
<tr>
<td><strong>Previous textbook evaluation experience</strong></td>
<td>No</td>
<td>Yes, selected books for new and existing courses.</td>
<td>Yes, served as materials development coordinator, curriculum development coordinator, and head of textbook writing project.</td>
</tr>
</tbody>
</table>

These participants were selected to ensure that they were of the same nationality and were all working within a similar teaching context. This was done to lessen the effects of culture as a variable within the study and to aid the subjects in their ability to associate with the teaching brief presented to them during the evaluation session. The textbook evaluated was the Intermediate level *Just Right* (Harmer 2004), comprising a teacher’s book, a student’s book, and a mini-grammar booklet containing transcripts of the listening activities. This coursebook was selected as being suitable for the teaching environment from which the participants were drawn, although it
was not in use in the participants’ institution, and therefore all three were equally unfamiliar with the contents.

**Training**

Another criticism sometimes levelled at think-aloud protocols is that not all participants are good at thinking aloud, or like doing it, any more than some teachers like to be observed, or are good at filling in questionnaires. With this in mind, prior to the start of the individual teachers’ evaluation sessions, the interlocutor asked each subject to engage in a warm-up activity to enable them to practise the technique of thinking aloud. The teachers were given three activities to choose from according to their personal preference: an anagram, the Tower of Hanoi puzzle, and a partially-completed jigsaw puzzle; all three chose the jigsaw. The interlocutor then presented the subjects with a brief containing details of a teaching situation similar to that within which each was working at the time to help contextualize the evaluation (see Figure 1 below). The participants were also reminded that one of the project goals was to gain insight into their thought processes and, therefore, it was important that they continued to verbalize throughout the session. The researchers refrained from providing any tips on evaluation.

![Figure 1](image1)

Please evaluate the suitability of the textbook provided for use with the group in the following scenario:

A group of 15–20 intermediate-level learners, predominately of Chinese nationality, who are learning basic English skills without any ESP (English for Special Purposes) or EAP (English for Academic Purposes) focus within a university setting.

**Analysis**

The subjects were audio- and video-taped while giving their think-aloud protocols. The session tapes were then transcribed and the transcripts distributed to the members of the research team for coding using a grounded approach. The team also watched the video tapes to explore coding ideas. Five categories were identified: sequence of evaluation, teacher preferences, use of terminology, methodological concerns, and flexibility in usage. Each member of the research group took responsibility for focusing in depth on one of these categories and analysing the data. The findings were then discussed by all and recategorized and reanalysed where necessary. The process continued until the salient features of the data were brought to light. The rest of this article will discuss two of the key findings of the study, and the implications of the project.

**Findings**

The sequence of evaluation

All the teachers began by looking at the brief provided, and then at the contents pages, and all at some stage looked at all the different components of the scheme: the student’s book, the teacher’s book, and the mini-grammar booklet. However, there was a clear difference among the three teachers in the route they followed while they were evaluating the textbook.

T1 started at Unit 1, but did not find an activity that she considered to be suitable for a first lesson (‘a warmer’) and so jumped to Unit 5. This was one manifestation of her greater concern for classroom logistics than for the
overall evaluation of the textbook. She then opened the teacher’s book and examined the description of the workbook and the methodological guidelines. This led her to look for the mini-grammar booklet, by flicking through the student’s book until she found it (tucked into a pocket in the back cover). Of the three participants, T1 seemed to be the least familiar with the commonalities of textbook layout. She then flicked through the mini-grammar booklet, and kept it to hand while she looked at various aspects of the student’s book: pronunciation, speaking (including the ‘activity bank’), functional language, and dictionary work. For each of these areas of language, she looked at how it was covered in three or four of the units, using the contents pages to find the appropriate sections. Finally, she returned to the teacher’s book and looked at the teaching notes in more detail.

T2 took a more direct approach, though it was not necessarily more time-efficient. He quickly found all the components of the book, looked at the front and back covers of the books and then flicked through both the student’s book and the teacher’s book before returning to the contents pages. He then looked through every page of the student’s book in turn, until he reached Unit 13 (of 14). At this stage he referred to the teacher’s book for Unit 1 and compared it to the student’s book. He repeated this process for six other units and then returned again to the beginning of the student’s book and flicked through both books once more, before giving an overall evaluation of the textbook package.

T3, on the other hand, adopted a more selective approach. He decided early on to focus on Unit 4, explaining that he expected the textbook to have ‘settled down’ by then. He looked at all the activities in this unit while referring to the other components of the course. Then he turned to the teacher’s book and compared it to the student’s book for Unit 4. He then looked at the methodology notes in the teacher’s book in more detail, before briefly looking through Unit 8, comparing the student’s book to the teaching notes.

Compared with T1’s ‘back and forth’ approach, T3’s sequence was very focused and time-efficient. T1 and T2 also showed some systematicity, but their evaluations appeared to be more superficial because they made only very brief comments on different parts of the various units, without achieving T3’s systematic approach at a deeper level. T3 knew what he was looking for in evaluating a textbook, and he knew where to find it. This difference could be vital because the great importance placed on textbooks in language education (e.g. Cunningsworth 1984; Hutchinson and Torres 1994) means that teachers need to know how to evaluate textbooks efficiently and effectively.

Teacher preferences

Not only did the three teachers follow different sequences in evaluating the textbook, but they also rated the usefulness of its main features differently.

T1 appeared to evaluate the textbook on the basis of whether it facilitated the smooth running of the lesson. This meant, in the first place, that she looked in the textbook for features that could help her ‘find out where she was’. She thus appreciated the uncluttered layout of the contents page and the clear
organization of the units and the teacher’s book. Of the tasks and activities featured in the textbook she judged negatively those that she believed may lead to problems in classroom management. For example, she did not like open-ended discussion tasks as she felt they were not particularly suited to certain groups of students, who, in her experience, are rather reluctant to interact with each other in English. She was also sceptical about the usefulness of split information tasks, which may lead to students ‘cheating’ and resorting to their native language. The ‘procedural’ aspect of the teaching activities is thus what the novice teacher seemed to be most concerned with. Hardly any attention was paid to the nature of the language featured in the textbook.

Rather than focusing exclusively on how the tasks and activities might work in the classroom, T2 seemed to place more emphasis on what learning outcomes they might have. For example, he preferred tasks where the student is involved as a prime actor rather than a passive spectator, as well as opportunities built into the textbook for ‘revisiting’ language introduced earlier in the book, as he felt that active involvement and recycling of language would be likely to lead to more thorough learning. He also showed some awareness of the linguistic content of the activities. Indeed, even though he did not always use the appropriate terminology to refer to linguistic phenomena, he identified and judged positively the focus on the collocational language of some activities. Throughout the evaluation, T2 seemed to be focused on what he perceived as the more immediate needs of the students. Hence, topical, up-to-the-minute themes, vocabulary, and functional language were rated particularly highly as being more likely to be relevant to the students’ everyday needs.

T3 seemed to comment on a wider range of issues than the other evaluators. In particular, many of his evaluations concerned the book’s linguistic rationale. For example, he was particularly impressed with the ‘lexical’ slant of the textbook, and the occurrence of ‘language in chunks’ in many activities. T3 was also less categorical and more flexible in expressing what he liked and did not like about the book. That is, he weighed up pros and cons of different features of the book, activities, and methodological options, and took great pains in accommodating other people’s (teachers’ as well as students’) needs and expectations. He was particularly concerned with the needs of less experienced teachers and the long-term academic needs of the students. Also, his evaluations often contained references to other textbooks he was familiar with.

The general impression is that the more teaching experience an evaluator has (which obviously also involves at least some experience in textbook evaluation) the more able he or she is to view a textbook with detachment, and take account of other users’ needs as well as his or her own. T1 equates the textbook with a script for lessons and prioritizes the teacher’s need for ‘survival’ when evaluating the book. T2 focuses on the student’s needs, although it is their immediate needs of functioning in an English-speaking environment that are of greatest concern. T3 manages to consider how the textbook fits into a long-term programme of preparation for academic study and how other teachers might relate to it. Also, whereas T1 seems to be looking for a textbook packed with activities to cut short the search for
supplementary materials, T2, and especially T3, put a high premium on the ‘jumping off’ opportunities that a textbook offers. What for T1 is a lifeline may seem like a straightjacket for the more experienced teachers.

The results of this study therefore appear to support the point made by Skierso (1991) that experienced teachers are better equipped than novices to adapt teaching materials for use according to their own individual style within particular teaching environments. Beginning teachers, by virtue of their inexperience in the classroom, require texts that include supplementary activities and detailed notes describing how they are to be used, as well as information regarding the pedagogical objectives that inform their design (ibid: 433).

**The implications of this study**

Expertise studies, like this one, are revealing from a theoretical point of view, providing insights of interest to the applied linguist. But arguably of more importance is their value in providing a research-informed basis for teacher training. For instance, those in training could be given the task of participating in similar types of materials evaluation studies, producing think-aloud protocols of the sessions, the transcripts of which they and their classmates could then analyse for purposes of determining and possibly improving their own evaluation styles. This suggestion for using data gathering methods to train teachers was proposed by Westerman (1991) following her study of teacher decision-making in order to ‘[promote] reflection for expert teachers as well as for novices’ (ibid: 303). An empirically-grounded method of teacher training such as this could go a long way towards helping student teachers to recognize their own textbook evaluation styles, as well as to learn more about the use of concurrent verbalization in research.

The LATEX study also points to the benefits of training teachers to take an in-depth approach to evaluation whereby the evaluator

\[
\text{[goes] beneath the publisher’s and author’s claims to look at \ldots the kind of language description, underlying assumptions about learning or values on which the materials are based or \ldots whether the materials seem likely to live up to the claims that are being made for them. (McGrath 2002: 27–8)}
\]

This certainly was the style employed by T3 during the study. Following this systematic method of evaluation seemed to enable him to take a more holistic stance in his assessment of the textbook and, in so doing, he was able to consider what lay behind the textbook’s design and aims.

Thus if we wish to train teachers to evaluate textbooks, we will benefit (one might argue) from knowing something about how experienced evaluators operate—what procedures they use and what issues they focus attention on. Given the limited generalizability of the results from this small-scale study, it is an overly-ambitious goal to develop any definitive statements with regard to training objectives. But further research into this area seems likely to reveal aspects of behaviour which would otherwise be ignored by trainers. It is true that the precise ways in which expertise studies can inform the trainer need to be thought about with care. Certainly it would be naïve to try and teach all the characteristics an expert possesses and think that thereby a
A step forward: investigating expertise in materials evaluation

References


The authors

The Language Teaching Expertise research group based within the Department of Linguistics and English Language at Lancaster University conducts original research every year into aspects of teaching expertise. The personnel of this group, made up of students and staff, changes yearly. Further details about this group can be found at http://www.ling.lancs.ac.uk/groups/latex/latex.htm.