Refinement of the English Language Teaching Textbook Evaluation Checklist

Vahid Nimehchisalem and Jayakaran Mukundan*

Department of Language and Humanities Education, Faculty of Educational Studies,
Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia

ABSTRACT

The English Language Teaching (ELT) Textbook Evaluation Checklist was developed in response to the need for a reliable, valid and practical instrument to evaluate English language teaching textbooks. The checklist was qualitatively developed by a review of the literature (Mukundan & Ahour, 2010; Mukundan *et al.*, 2011a) and was refined through qualitative (Mukundan *et al.*, 2011b) and quantitative (Mukundan & Nimehchisalem, 2012a) methods. As the validation test results of the checklist (Mukundan & Nimehchisalem, 2012b; Nimehchisalem & Mukundan, 2013) indicated, it could be refined further to improve its validity, reliability and practicality. The present study discusses the modifications made to the checklist following the comments of a panel of experts (n=3), who were sent a copy of the old version of the checklist. They commented on the comprehensiveness, importance and clarity of the domains and items of the checklist independently. The qualitative method was used to collect and analyse the data. The checklist was refined based on the experts' comments; problematic items were removed or revised and a scoring guide was added to it. The refined instrument is more economical than its previous version, and yet further research is required to test its validity empirically.

INTRODUCTION

Textbooks are evaluated for two main reasons: selection or adaptation (Sheldon, 1988). Sometimes programme developers or language teachers need to evaluate textbooks so that they can make wise decisions in selecting the most suitable book for a language course. Textbook

selection can remarkably influence the process of learning and teaching as students and teachers continuously make references to them (McGrath, 2002) and programme developers design syllabi around them (Garinger, 2002). Textbook evaluation can also help the reflective teacher make necessary changes to a textbook by examining its weaknesses or strengths. Teachers' involvement in this type of evaluation can help them develop professionally (Mukundan, 2010). However, in practice, teachers often prefer to use books that have attractive covers or use best-selling books blindly and merely because they are used by others (Tomlinson, 2010). It is possible to evaluate textbooks in two ways, implicitly or explicitly. Language teachers often evaluate textbooks implicitly, which involves an impressionistic judgment of the material through a quick glance at it with a picture of the target learners in mind. One may find it easy to rely on the impressionistic judgment of an experienced teacher to test the suitability of a textbook. However, less experienced teachers, who have not yet developed a principled set of evaluative criteria in their minds, will find it hard to come up with valid judgments. In such cases, a better choice would be using the explicit evaluation method in which a clear set of criteria is referred to by the evaluator to test the usefulness of the material. Instruments that present such explicit sets of criteria are called checklists. They can help the evaluator make an explicit and comprehensive assessment of the quality of the textbook in question. Checklists can reduce the subjectivity of evaluators' impressionistic judgment, contributing to the reliability of the evaluation.

A review of the available checklists indicates that in spite of their importance, these instruments are often not tested for their validity or reliability. Admittedly, there are checklists that have been validated, but they are rarely tested for their practicality. Some of these checklists are laden with specialised terminology that can be discouraging for language instructors with little theoretical knowledge of ELT. There are also some checklists whose developers report temptingly high reliability indices. However, in practice, most language instructors will avoid these because they are highly specialised and/or uneconomical. The ELT Textbook Evaluation Checklist (reported here) was developed to provide language instructors and researchers with a valid, reliable as well as practical instrument to evaluate English teaching textbooks. The checklist was developed using both qualitative and quantitative methods of analysis. The next section briefly reviews the literature on ELT material evaluation and gives a summary of the procedure of developing the present checklist.

LITERATURE REVIEW

Different scholars have suggested various procedures and criteria for evaluating ELT materials. Tomlinson (1998) provides an extensive list of principles to be considered in material development and evaluation.

He proposes that good English language learning-teaching materials should:

1. achieve impact (by having novelty and variety and by being attractive and appealing),
2. help learners feel at ease and overcome anxiety (by providing plenty of white space and providing comprehensible input, i+1, Krashen, 1988),
3. be relevant to the learners' needs,
4. promote learner self-investment (by involving them in projects and creating their own resources),
5. present authentic language,
6. draw learners' attention to linguistic features of the input (which does not necessarily mean an explicit presentation of grammar),
7. promote the use of the target language to achieve communicative purposes (e.g., through information gap activities,
8. consider the fact that positive effects of instruction are often delayed,
9. consider learners' varying learning styles and affective attitudes,
10. permit a silent period at the beginning of instruction,
11. encourage intellectual, aesthetic and emotional involvement that stimulates both right and left brain hemispheres (through activities like singing a song),
12. not rely on controlled practice (since it is retained only in short-term memory), and
13. provide opportunities for outcome feedback (by activities that encourage them to check their language achievement).

According to Stevick (1971), textbooks could be evaluated based on three qualities (strength, lightness and transparency), three dimensions (linguistic, social and topical) as well as four components (occasion for use, sample of language use, lexical exploration and exploration of structural relationships). Brown (1995) suggests evaluating materials in terms of their background (e.g. the author's credentials), fitness to the target curriculum, physical characteristics, logistical characteristics (e.g. price and availability) and teachability (e.g. annotations to help teachers explain and plan activities). For Ansari and Babaii (2002), approach, content presentation, physical make-up and administrative concerns constitute the universal features that are the bases on which a textbook should be evaluated. Bell and Gower (1998) emphasise that the language presented in an English textbook should be natural and realistic. They contend that a textbook with authentic examples of language commonly used in the real world can motivate the learner.

Some other dimensions that are often considered in evaluating textbooks include the learner's role, the teacher's role and the role of instructional materials (Richards & Rodgers, 1986), as well as the learning-teaching context and the background of the learners (Byrd, 2001; Sheldon, 1988; Skierso, 1991). Another crucial factor is the relevance of learning content in the ELT materials in which the format and orientation of activities should be geared to learners' preferences and developmental needs (Tudor, 1992).

Although the available checklists have been developed for different learning-teaching situations, they usually share almost the same characteristics and evaluative criteria. Features such as aim, layout, methodology, organisation, language skills (speaking, listening etc.), sub-skills (grammar, vocabulary etc.) and functions can be frequently seen in most of the commonly used checklists like Cunningsworth (1995), Cunningsworth and Kusel (1991), Harmer (1991), Skierso (1991) and Ur (1996).

The content of a checklist should be relevant and explicit. The items in a checklist should be worded in such a way that the concepts are defined clearly and unambiguously. In this respect, Littlejohn (1998) warns checklist developers about including items that involve "general, impressionistic judgements … rather than examining in depth what the materials contain" (p.191). Ur (1996) recommends that developers avoid wording items in such a way that they are limited to a specific methodology or approach. According to Littlejohn (1998), it is important to avoid items like "up-to-date methodology of L2 teaching" (as in Williams, 1983, p.252). A teaching method that is up-to-date may not be necessarily the best available method for the target learning-teaching situation.

Furthermore, concepts that are hard to quantify should be avoided. For example, it is very hard to score an item that asks whether the material is "based on a contrastive analysis of English and the L1 sound system" (Williams, 1983, p.255).

Checklists are used by language teachers, who often cannot spare time to make a contrastive analysis of the first and target languages merely to be able to accurately respond to an item in a checklist. Also, teachers would find it very hard to quantify an item like "Balanced distribution: To what extent is there an even distribution of grammatical and vocabulary material among the chapters?" (as in Skierso, 1991, p.447). An analysis of such distribution patterns would require software like Wordsmith (Scott, 1999) or Retrotext-E (Mukundan, 2010). A human numerator would find it very challenging to analyse the distribution patterns of the new words throughout a textbook manually.

According to Tomlinson (2003), there are certain pitfalls that should be avoided by developers of textbook checklists, including (1) confusion over evaluation and analysis questions, (2) multiple questions in one item, (3) extended, unclear and unanswerable items, (4) dogmatic questions and finally (5) items that may be interpreted in different ways by various evaluators. In their review of 48 available checklists in the literature, Mukundan and Ahour (2010) show how checklist developers commonly commit the errors mentioned by Tomlinson (2003).

PREVIOUS RESEARCH ON THE ELT TEXTBOOK EVALUATION CHECKLIST

The ELT Textbook Evaluation Checklist (referred to as 'the checklist' henceforth) has undergone a number of recursive

validation and refinement stages that will be presented in this section. The checklist was first designed based on the review of the related literature on ELT textbook evaluation (Mukundan & Ahour, 2010). Based on this review, a prototype was developed (Mukundan *et al.*, 2011a). It consisted of two main domains, namely 'general attributes' (5 sub-domains and 11 items) and 'learning-teaching content' (9 sub-domains and 27 items). The sub-domains in the first section included 'the book in relation to syllabus and curriculum', 'methodology', 'suitability to learners', 'physical and utilitarian attributes' and 'supplementary materials'. On the other hand, 'general content' (i.e. task quality, cultural sensitivity as well as linguistic and situational realism), 'listening', 'speaking', 'reading', 'writing', 'vocabulary', 'grammar', 'pronunciation' and 'exercises' constituted the sub-domains of the second section. The checklist followed a five-point Likert style scale, in which zero signified 'never true' and four 'always true'.

The prototype was further refined through qualitative and quantitative methods. Six experts participated in a focus group study to comment on the clarity, comprehensiveness and importance of the checklist items, sub-domains and domains. Based on the focus group's feedback, 2 items were rephrased and 14 more were added to the checklist (Mukundan *et al.*, 2011b). The focus group study was very useful. Most importantly, it helped the developers revise some of the ambiguous items that would otherwise have reduced the reliability of the instrument.

Subsequent to the focus group study, the refined checklist was given to a group of ELT experts (n=207) who stated their views on the checklist. They indicated which items needed to be reworded, removed or added. They also determined the importance level of each item by marking a five-point Likert scale from zero, signifying 'unimportant', to four, signifying 'very important'. The method that was followed was almost the same as the one followed by Akbari *et al.* (2010) or Cid *et al.* (2009). Based on the results of factor analysis, two items were removed from the checklist. A discussion of the findings of this survey is presented in Mukundan and Nimehchisalem (2012a).

After the focus group study and the survey, the checklist turned into a 50-item instrument. At this point, two other studies were simultaneously conducted to validate the checklist. One of the studies involved a survey of a group of English language teachers' views on the usefulness (validity, reliability, impact and practicality) of the checklist. For this purpose, a group of teachers (n=82) with bachelor degrees and a minimum of 2 years of teaching experience used the checklist to evaluate a textbook. They were then given a questionnaire to evaluate the checklist itself. The findings showed that more than three out of four teachers (78%) regarded the checklist as highly useful (Nimehchisalem & Mukundan, 2013). Additionally, some of the teachers provided comments that helped the developers in the further refinement of the checklist. The study provided some promising feedback on the usefulness of

the checklist, but its main limitation was its reliance on the 'perceptions' of a group of respondents rather than on a panel of experts' evaluation of the checklist. It could not be concluded that the results of this survey were really able to take all of such elusive facets as validity, reliability, impact and practicality into account.

Therefore, another study was conducted which involved a more empirical evaluation of the validity, reliability and economy of the checklist. In this study, two ELT experts, who were experienced English language teachers with PhD degrees in TESL, evaluated the same textbook using the checklist. The correlation between the two raters' results (the inter-rater reliability) proved to be quite high (r=.88). The experts also used another checklist with proven validity (i.e. Skierso, 1991) to evaluate the same textbook. A high correlation (r=.77) was reported between the results of the two checklists. This would mean that the checklist results indicated a high level of concurrent validity. Additionally, the checklist turned out to be more economical than the Skierso checklist. It helped the two experts evaluate the textbook in 9.5 minutes, which was almost three times shorter than the time they spent using the Skierso checklist for the same purpose. Finally, the experts' responses to a questionnaire indicated they were highly satisfied with the checklist (Mukundan & Nimehchisalem, 2012b).

As the reviewed literature on the ELT Textbook Evaluation Checklist indicates, the instrument has been developed and tested rigorously and positive results have been obtained on its validity, reliability, impact and practicality. However, further in-depth research was necessary to shed light on its potential problems and help its developers increase its usefulness. The present study was proposed to address this objective.

OBJECTIVE AND RESEARCH QUESTIONS

This study aimed at validating the ELT Textbook Evaluation Checklist with the help of a panel of ELT experts. The following research questions were posed to meet the objective of the project:

1. How do the experts evaluate the ELT Textbook Evaluation Checklist regarding the comprehensiveness of its items, sub-domains and domains?
2. How do the experts evaluate the checklist regarding the importance of its items, sub-domains and domains?
3. How do the experts evaluate the checklist in reference to the clarity of its items?

METHOD

This research was an attempt to further improve the validity (research questions 1 & 2) and reliability (research question 3) of the checklist. Qualitative method was used to collect and analyse data. The panel of experts that were consulted included three experienced ELT experts (one female and two males). They were all experienced ELT practitioners (with a minimum experience of 15 years) and all held PhD

degrees in Teaching English as a Second Language (TESL). They were individually given a copy of the latest version of the checklist with a cover letter indicating the objective of the research and asking them for their critical feedback. They were particularly requested to review the checklist and write down their comments on the comprehensiveness, importance and clarity of its items. All the experts returned their comments in less than a week.

Using a panel of experts' judgment, or qualitative expert reviews, is a common way of validating instruments (Wynd, Schmidt, & Schaefer, 2003). Also known as the Delphi technique, the method involves obtaining an expert panel's independent evaluations over repeated rounds until consent is reached (Armstrong, 2001). After each round, summaries of the anonymous judgments are provided for the experts (Rowe & Wright, 2001). Delphi is based on the assumption that group judgment has higher validity than individual and/or unstructured group judgments and provides more accurate forecasts than unstructured groups (Rowe & Wright, 1999). In this study, the three experts provided consistent judgments about the validity of the instrument, which rendered a second round redundant.

RESULTS AND DISCUSSION

The study sought to elicit feedback from the experts on the comprehensiveness, importance and clarity of the checklist items. The experts added some items that they believed would make the checklist more comprehensive (research question 1). They removed some that they considered unimportant (research question 2), while they modified and reworded the items that were not clear (research question 3). This section presents and discusses the changes made to the checklist based on the experts' comments in three sections, each covering a separate research question. When references are made to the old version of the checklist (Appendix A), item numbers are followed by the word 'Old' (e.g. item 1 Old); otherwise, the item appears in the new version of the checklist (Appendix B).

Added Items

Some items were added to the checklist based on the experts' comments. These included the flexibility of the checklist, textbook illustrations, supplementary materials (like tests and workbooks), language functions, learner attitudes and register of listening contexts. A score interpretation guide was also added to the checklist. Some of these comments were considered while others were disregarded, as discussed in this section.

There were comments concerning the flexibility of the checklist. The researchers were recommended to add a part to the checklist to enable the evaluators to disregard any of the items that they considered irrelevant to their present teaching situation. Some well-known checklists permit the evaluator to ignore an item that lacks relevance to particular learning-teaching contexts. The Skierso checklist (1991), for instance, has a 'not

applicable' column next to each item. Should an evaluator find an item irrelevant for the teaching-learning context in question, s/he may mark it as 'not applicable' and disregard it. In order to make the present checklist more flexible, a 'not applicable' column was added next to each sub-domain. A note was also added to the beginning of the checklist which instructed evaluators to disregard the sections that they found irrelevant to their present context.

The second expert suggested adding the item, *Illustrations are sufficient*, to Part D. This comment was, however, disregarded since it was covered by item 9, *It indicates efficient use of text and visuals*. In part E (Efficient layout of supplementary materials), the second and third experts both suggested adding items on tests, quizzes and workbooks. Adding three items to this part would reduce the economy of the checklist. Instead, item 14 Old, *The book is supported efficiently by essentials like audio-materials*, was modified to *The book is supported by suitable materials, like a workbook, audio, or multimedia* (item 10). A new item was also added to cover assessment, *The book is supported by other materials like review and test units* (item 11). One of the experts suggested adding an item on external links and references to internet sources and the like. This recommendation was, however, ignored since it would not be fair to penalise a book merely for not having made references to the internet.

The first expert suggested adding an item on language functions to part F

(General content) e.g. *Book covers the essential functions learners need to express their ideas in real life communication*. This item was not, however, added to the new version of the checklist for a number of reasons. First, the item would not be suitable for textbooks that do not follow a functional notional syllabus. Additionally, the first item of the checklist, *It matches to the specifications of the syllabus*, covered the same concept. Adding similar items to the checklist would also reduce its economy.

The first expert also suggested adding an item on the learners' attitudes towards the book. He recommended an item like *The students have a positive attitude toward the book*. The item was not added to the checklist since it would divert the instrument from its main objective, which is enabling teachers and other experts to evaluate the usefulness and suitability of ELT textbooks rather than measuring learners' attitudes or perceptions of textbooks.

Following the third expert's comment on part G (Listening), item 20, *Various listening contexts such as formal vs. informal contexts are considered*, was added to emphasise the variety of listening activities. Adding such an item was appropriate since it focused on the crucial area of style and register.

Finally, an interpretation guide was also added to the end of the checklist. To evaluate a textbook, the evaluator would read each item, assign a value of 0 to 4 for each, add up the scores to calculate the total score, and then divide this score by the total number of the items (i.e. 39) to

achieve the mean score. If the mean score ranged between 0 and 0.80, it would have 'negligible usefulness' to the target group of learners. A mean, ranging between 0.81 and 1.60, would indicate 'low usefulness'. The third level ranged between 1.61 and 2.80, which would indicate 'moderate usefulness'. The next level, ranging from 2.81 to 3.60, would show 'high usefulness'. The final level that fell between 3.61 and 4.00 would indicate a 'very high level of usefulness'. The levels were defined based on Guilford's (1973) rule of thumb:

    <20%   = negligible
    20%-40% = low
    40%-70% = moderate
    70%-90% = high
    >90%   = very high

According to this rule of thumb, a value of <20% is regarded as 'negligible', 20%-40% as 'low', 40%-70% as 'moderate', 70%-90% as 'high' and >90% as 'very high'.

Removed Items

This section will present the items that were deleted from the checklist. The items were removed generally for two main reasons. Most of these items overlapped with other similar items in the checklist. Other items were removed since they would make the checklist suitable only for particular types of English textbooks, making it less generic.

Item 5 Old, *It is compatible with the socio-economic context*, was deleted because the concept of learners' socio-economic context would be covered by item 7, *It is appropriately priced*. Items

6 Old, *It is culturally accessible to the learners*, and 18 Old, *Cultural sensitivities have been considered*, were also removed. According to the third expert, these items had already been covered by item 4 Old, *It is compatible with the background knowledge and level of students*. Item 30 Old, *Activities motivate students to talk*, in part H (Speaking) was removed since item 21, *Activities are developed to initiate meaningful communication*, would cover it.

Item 20 Old, *The situations created in the dialogues sound natural and real*, was removed since it would not be fair to regard an English textbook as unsuitable for lack of situational realism in it. In addition, linguistic realism (item 19 Old) to some extent overlapped with this item.

Item 23 Old, *The book contains fun elements*, was removed. Admittedly, engaging language students can definitely increase their motivation and make them more successful learners. However, it is the teacher and, in part, the students, who can make the textbook fun. In this respect, Pulverness (1999) contends that it is the teacher who reanimates the dead text of the book for the learners. As the second expert also noted, different cultures may have different conceptions of 'fun'. In addition, if the textbook has 'varied' (item 14) or 'interesting' (item 27) tasks, it will make it easy for the teacher to make its use fun.

Items 26 Old, *Tasks are efficiently graded according to complexity*, and 31 Old, *Texts are graded*, were removed based on the first expert's feedback. He contended:

Some items imply that 'gradation' is an essential criterion of suitability. However, gradation is a characteristic of linguistic or grammatical syllabi, and may be done away with by analytic syllabi or process-oriented syllabi. In addition, according to Ellis (2003), "there is no simple algorithmic procedure for grading tasks in terms of their inherent characteristics" (p. 122).

In fact, items 5, *It is compatible with background knowledge and level of students*, and 13, *Tasks move from simple to complex*, already covered the concept of difficulty, making items 26 and 31 redundant.

Finally, item 35 Old, *Models are provided for different genres*, was also removed. The item would be more suitable for English textbooks following a genre- or text-based approach to teaching of ESL writing. The approach is more appropriate for courses following English for academic purposes. Having an item that promoted this approach would unfairly under-rate the textbooks that followed a different approach.

Modified Items

The panel of experts also suggested modifying some of the items for a number of reasons. As in the case of items that needed to be removed, some items lacked unidimensionality and focused on more than one sub-construct. There were also items that overlapped with others. It is common practice to have more than one item to test the same sub-construct. In the development of the present checklist, however, we tried to avoid overlapping

items to keep the final checklist as concise as possible. The experts also highlighted some ambiguous items that were too broad and could confuse the evaluator. Some other items were modified to make the checklist more economical.

As the third expert pointed out, item 2 Old, *The activities can be exploited fully and can embrace the various methodologies in ELT*, was testing two separate sub-constructs. The second part of this item, *...and can embrace the various methodologies in ELT*, was, in fact, the same as item 3 Old, *Activities can work well with methodologies in ELT*. This part of the item was, therefore, removed.

As it was also observed by the first expert, item 4 Old, *It is compatible with the background knowledge and level of students*, focused on two separate issues. In this respect, Tomlinson (2003) warns checklist developers of including multiple questions in a single item. The learners' background knowledge includes their world knowledge, but their level encompasses their linguistic knowledge. Thus, the item was broken into two separate items.

Another item that needed to be modified was item 3 Old, *Activities can work well with methodologies in ELT* since, according to the second expert, it was too broad. In her words, "Do we expect a book to cater for or accommodate various methodologies? Then the evaluator needs to have access to your operational definition of methodology." The third item was, therefore revised as, *Activities can work well in most classroom situations* (item 4) to avoid the ambiguity that the

word 'methodology' could create. One of the experts commented on the word 'needs' in item 7 Old, *It addresses the needs of the learners*. As he argued, "The word 'needs' is too general a concept. It could include a variety of needs such as linguistic needs, functional needs, communicative needs etc.; moreover, different students may have different needs." The item was, therefore, rephrased to *It addresses learning targets* (item 6) to avoid ambiguity.

Item 11 Old, *It is cost-effective*, was rephrased to *It is appropriately priced*. As the third expert mentioned, in some cases free English textbooks may be provided for the students. The revised item can cover such situations. In part F (General content), item 17 Old, *Task objectives are achievable*, was reworded more clearly as *Tasks support teaching objectives*. Item 21 Old, *The material is up-to-date*, was also modified, based on the second expert's comment. According to her, the term 'up-to-date' is vague and "can have different connotations. It may relate to the date of publication, the content, the topics, or the like". The item was reworded as *The content is fairly recent* to avoid confusion. Another item that was modified was item 22 Old, *It covers a variety of topics from different fields*. The idea of variety cannot be neglected since it makes the textbook more interesting. However, in the case of some textbooks, it would be impossible to have a variety of topics from different fields. Therefore, the focus of the item was shifted from the topics to the tasks. It was reworded as *Tasks are varied* (item 14).

As mentioned by two of the experts, factors like durability, size, or printing quality are important, but not as essential as other items like the compatibility of the book to the background knowledge and level of the students (item 4 Old). Therefore, items 10 Old, *It is durable*, 12 Old, *Its size is appropriate*, and 13 Old, *The printing quality is high*, were all rephrased more generally as *Overall, the book has a nice feel* (item 2). Collapsing these items would also improve the economy of the checklist. Likewise, in part H (Speaking), item 29 Old, *Activities are balanced between individual work, pair work and group work*, was rephrased as *Individual, pair and group work are given equal emphasis* (item 22). The reworded item was shorter and easier to understand.

CONCLUSION

This paper summarises the previous research conducted to develop and evaluate The ELT Textbook Evaluation Checklist. It also presents the findings of a study that investigated a panel of experts' evaluation of the validity and reliability of the checklist. This study helped the developers gain further in-depth feedback from the prospective users of the checklist regarding its usefulness. Based on this feedback, the checklist was revised. The new version of the checklist (see Appendix B) is more concise than the previous version. It has 39 items (with 303 running words) while the previous version contained 50 (with 361 running words). It also makes it more convenient for the evaluator to rate and

interpret textbooks with the help of its interpretation table that was missing in its predecessor.

This study has useful implications for researchers who are interested in developing similar instruments. As the results indicated, there were many items in the checklist that were ambiguous or overlapped with others. Such issues could affect the reliability and validity of the instrument, but surprisingly they had been ignored in the previous studies through which the checklist had been developed. This leads us to the important conclusion that it is not sufficient to validate an instrument based on the perceptions of a group of experts or even the evaluation of a group of users. The findings have led us to believe that the critical feedback of a panel of experts can shed light on many areas of an instrument that may be ignored in more positivistic research methods.

The study emphasises the role of textbook evaluation checklists in ELT research and practice. A checklist is a useful tool that can be used for research purposes; that is, "measuring what might be expected to occur; and as a means of raising awareness among designers, i.e. pointing out what perhaps ought to occur" (Jones, 1993, p.457). As a research tool, a checklist can facilitate and encourage self-directed teacher research, which has been underlined in the literature (Elliot, 1981; Roberts, 1993). It can be helpful in 'pre-use', 'in-use' and/or 'post-use' ELT material evaluation (Cunningsworth, 1995; Ellis, 1997). In pre-use (or predictive) evaluation of textbooks, the checklist can help teachers

to select more useful and suitable textbooks for their language classes by examining the prospective performance in their present learning-teaching context. Teachers can use checklists to diagnose the strengths and/or weaknesses of the textbooks that they are using in their classrooms. Finally, checklists can also be used in post-use (or retrospective) evaluation of textbooks that involves a reflection on the quality of a textbook that has been taught/used in a language classroom. The present checklist can be used by programme evaluators, researchers and other experts to evaluate a textbook after it has been used. The probable shortcomings of the textbook can specifically be diagnosed with the help of the checklist and adaptations can be made to optimise the usefulness of the textbook.

However, any instrument should be used with caution as testing situations vary. For instance, the checklist will definitely require some modifications before it can be adequately used to evaluate junior level English textbooks for children. Additionally, the checklist cannot be used for evaluating ELT software packages unless certain criteria are added to address the technical properties of the software. It is common to test checklists by using them to evaluate a wide range and number of materials (Jones, 1993). A study in which the checklist is used to evaluate several textbooks may help further improve its validity.

REFERENCES

Akbari, R., Behzadpour, F., & Dadvand, B. (2010). Development of English language teaching reflection inventory. *System, 38*(2), 211–227.

Ansary, H., & Babaii, E. (2002). Universal characteristics of EFL/ESL textbooks: A step towards systematic textbook evaluation. *The Internet TESL Journal VIII (2)* Retrieved March 1, 2012, from http://iteslj.org/Articles/Ansary-Textbooks/

Armstrong, J. S. (Ed.). (2001). *Principles of forecasting: A handbook of researchers and practitioners*. Boston: Kluwer Academic Publishers.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice.* Oxford, UK: Oxford University Press.

Bell, J., & Gower, R. (1998). Writing course materials for the world: A great compromise. In B. Tomlinson (Ed.). *Material development in language teaching* (pp. 116–129). Cambridge: Cambridge University Press.

Brown, J. D. (1995). *The elements of language curriculum*. Boston: Heinle & Heinle Publishers.

Byrd, P. (2001). Textbooks: Evaluation for selection and analysis for implementation. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language*. (3rd Ed.) (pp. 415–427). US: Heinle & Heinle, Thomson Learning Inc.

Cid, E., Granena, G., & Tragant, E. (2009). Constructing and validating the foreign language attitudes and goals survey (FLAGS). *System, 37*(3), 496–513.

Cunningsworth, A. (1995). *Choosing your coursebook*. Oxford: Heinemann.

Cunningsworth, A., & Kusel, P. (1991). Evaluating teachers' guides. *ELT Journal, 45*(2), 128–139.

Elliot, J. (1981). *Action research: A framework for self-evaluation.* Schools Council.

Ellis, R. (1997). *SLA research and language teaching.* Oxford: Oxford University Press.

Ellis, R. (2003). *Task-based language learning and teaching.* Oxford: Oxford University Press.

Garinger, D. (2002). Textbook selection for the ESL classroom. *Center for Applied Linguistics Digest*. Retrieved from http://www.goenglish.rmutl.ac.th/wp-content/uploads/2013/03/Material-Design.pdf

Guilford, J. P. (1973). *Fundamental statistics in psychology and education* (5nd ed.). New York: McGraw-Hill.

Harmer, J. (1998). *How to teach English?* Harlow, Essex: Addison Wesley Longman.

Jones, F. R. (1993). Beyond the fringe: A framework for assessing teach-yourself materials for ab initio English speaking learners. *System, 21*(4), 453–469.

Krashen, S. D. (1988). *Second language acquisition and second language learning.* Herts: Prentice-Hall International.

Littlejohn, A. (1998). The analysis of language teaching materials: Inside the Trojan Horse. In B. Tomlinson (Ed.), *Materials development in language teaching*. (pp.190–216). Cambridge: Cambridge University Press.

McGrath, I. (2002). *Material evaluation and design for language teaching*. Edinburgh: Edinburgh University Press.

Mukundan, J. (2009). *ESL textbook evaluation: A composite framework*. Köln, Germany: Lambert Academic Publishing.

Mukundan, J. (2010). Retrotext-E 1.0: The beginnings of computer-based ELT textbook evaluation. *Advances in Language and Literary Studies*, *1*(2), 270–280.

Mukundan, J., & Ahour, T. (2010). A review of textbook evaluation checklists across four decades (1970-2008). In B. Tomlinson, & H. Masuhara (Eds.), *Research for materials development in language learning: Evidence for best practice* (pp. 336-352). London: Continuum.

Mukundan, J., Hajimohammadi, R., & Nimehchisalem, V. (2011a). Developing an English language textbook evaluation checklist. *Contemporary Issues in Education Research, 4*(6), 21–27.

Mukundan, J., Nimehchisalem, V., & Hajimohammadi, R. (2011b). Developing an English language textbook evaluation checklist: A focus group study. *International Journal of Humanities and Social Science, 1*(12), 100–105.

Mukundan, J., & Nimehchisalem, V. (2012a). Evaluative criteria of an English language textbook evaluation checklist. *Journal of Language Teaching Research*, *3*(6), 1128–1134.

Mukundan, J., & Nimehchisalem, V. (2012b). Evaluating the validity and economy of the English language teaching textbook evaluation checklist. *World Applied Sciences Journal, 20*(3), 458-463.

Nimehchisalem, V., & Mukundan, J. (2013). Usefulness of the English language teaching textbook evaluation checklist. *Pertanika Journal of Social Sciences and Humanities (JSSH), 21*(2), 797–816.

Pulverness, A. C. (1999). Context or pretext? Culture, content and the coursebook. *Folio, 5*(2), 5–10.

Richards, J., & Rodgers, T. (1986). *Approaches and methods in language teaching.* Cambridge: Cambridge University Press.

Roberts, J. R. (1993). Evaluating the impacts of teacher research. *System, 21*(1), 1–19.

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*, *15*(4), 353–375.

Rowe, G., & Wright, G. (2001): Expert opinions in forecasting: Role of the Delphi technique. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook of researchers and practitioners* (pp. 125-144). Boston: Kluwer Academic Publishers.

Scott, M. (1999). *WordSmith Tools, Version 4.0*. Oxford: Oxford University Press.

Sheldon, L. E. (1988). Evaluating ELT textbooks and materials. *ELT Journal, 42*(4): 237–246.

Skierso, A. (1991). Textbook selection and evaluation. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (pp. 432–453). (2nd Ed.). Boston: Heinle & Heinle Publishers.

Stevick, E. W. (1971). *Adapting and writing language lessons.* Washington, DC: Government Printing Office.

Tomlinson, B. (Ed.). (1998). *Materials development in language teaching*. Cambridge: Cambridge University Press.

Tomlinson, B. (2003). Materials evaluation. In B. Tomlinson, (Ed.), *Developing materials for language teaching* (pp. 15–36). London: Continuum.

Tomlinson, B. (2010). Principles of effective material development. In N. Harwood (Ed.), *English language teaching materials: Theory and practice*. New York: Cambridge University Press.

Tudor, I. (1992). Learner-centredness in language teaching: Finding the right balance. *System, 20*(1), 31–44.

Ur, P. (1996). *A course in language teaching: Practice and theory*. Cambridge: Cambridge University Press.

Williams, D. (1983). Developing criteria for textbook evaluation. *ELT Journal, 37*(3), 251–255.

Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two quantitative approaches for estimating content validity. *Western Journal of Nursing Research*, *25*(5), 508–518.

APPENDIX A

English Language Teaching Textbook Evaluation Checklist (Old Version)

Instructions

Read the items in the checklist and in the column opposite the items indicate the level to which they agree with each statement by marking 0 to 4:

0 = NEVER TRUE

1 = RARELY TRUE

2 = SOMETIMES TRUE

3 = OFTEN TRUE

4 = ALWAYS TRUE

I. General attributes A. The book is in relation with the syllabus and curriculum					
1. It matches the specifications of the syllabus.	⓪	①	②	③	④
B. Methodology					
2. The activities can be exploited fully and can embrace the various methodologies in ELT.	⓪	①	②	③	④
3. The activities can work well with methodologies in ELT.	⓪	①	②	③	④
C. Suitability to learners					
4. It is compatible with background knowledge and level of students.	⓪	①	②	③	④
5. It is compatible with the socio-economic context.	⓪	①	②	③	④
6. It is culturally accessible to the learners.	⓪	①	②	③	④
7. It is compatible with the needs of the learners.	⓪	①	②	③	④
D. Physical and utilitarian attributes					
8. Its layout is attractive.	⓪	①	②	③	④
9. It indicates efficient use of text and visuals.	⓪	①	②	③	④
10. It is durable.	⓪	①	②	③	④
11. It is cost-effective.	⓪	①	②	③	④
12. Its size is appropriate.	⓪	①	②	③	④
13. The printing quality is high.	⓪	①	②	③	④
E. Efficient layout of supplementary materials					
14. The book is supported efficiently by essentials like audio-materials.	⓪	①	②	③	④
15. There is a teacher's guide to aid the teacher.	⓪	①	②	③	④

II. Learning-teaching content					
A. General					
16. Tasks move from simple to complex.	⓪	①	②	③	④
17. Task objectives are achievable.	⓪	①	②	③	④
18. Cultural sensitivities have been considered.	⓪	①	②	③	④
19. The language in the textbook is natural and real.	⓪	①	②	③	④
20. The situations created in the dialogues sound natural and real.	⓪	①	②	③	④
21. The material is up-to-date.	⓪	①	②	③	④
22. It covers a variety of topics from different fields.	⓪	①	②	③	④
23. The book contains fun elements.	⓪	①	②	③	④
B. Listening					
24. The book has appropriate listening tasks with well-defined goals.	⓪	①	②	③	④
25. Instructions are clear.	⓪	①	②	③	④
26. Tasks are efficiently graded according to complexity.	⓪	①	②	③	④
27. Tasks are authentic or close to real language situations.	⓪	①	②	③	④
C. Speaking					
28. Activities are developed to initiate meaningful communication.	⓪	①	②	③	④
29. Activities are balanced between individual response, pair work and group work.	⓪	①	②	③	④
30. Activities motivate students to talk.	⓪	①	②	③	④
D. Reading					
31. Texts are graded.	⓪	①	②	③	④
32. Length is appropriate.	⓪	①	②	③	④
33. Texts are interesting.	⓪	①	②	③	④
E. Writing					
34. Tasks have achievable goals and take into consideration learner capabilities.	⓪	①	②	③	④
35. Models are provided for different genres.	⓪	①	②	③	④
36. Tasks are interesting.	⓪	①	②	③	④
F. Vocabulary					
37. The load (number of new words in each lesson) is appropriate to the level.	⓪	①	②	③	④
38. There is a good distribution (simple to complex) of vocabulary load across chapters and the whole book.	⓪	①	②	③	④
39. Words are efficiently repeated and recycled across the book.	⓪	①	②	③	④
40. Words are contextualized.	⓪	①	②	③	④

G. Grammar	⓪	①	②	③	④
41. The amount of grammar is achievable.	⓪	①	②	③	④
42. The grammar is contextualized.	⓪	①	②	③	④
43. Examples are interesting.	⓪	①	②	③	④
44. Grammar is introduced explicitly.	⓪	①	②	③	④
45. Grammar is reworked implicitly throughout the book.	⓪	①	②	③	④
H. Pronunciation					
46. It is contextualized.	⓪	①	②	③	④
47. It is easy to learn.	⓪	①	②	③	④
I. Exercises					
48. They have clear instructions.	⓪	①	②	③	④
49. They are adequate.	⓪	①	②	③	④
50. They help students who are under/over-achievers.	⓪	①	②	③	④

APPENDIX B

English Language Teaching Textbook Evaluation Checklist (New Version)

Instructions

Read the items in the checklist and in the column opposite the items indicate the level to which they agree with each statement by marking 0 to 4:

0 = NEVER TRUE	3 = OFTEN TRUE
1 = RARELY TRUE	4 = ALWAYS TRUE
2 = SOMETIMES TRUE	NA= NOT APPLICABLE

NB: There may be one or more sections that you may not find applicable to your teaching context. In such cases, check the box in the 'NA' (or, NOT APPLICABLE) column and disregard the section in evaluating the textbook.

NA	I. General attributes					
☐	A. The book in relation to syllabus and curriculum					
	1. It matches the specifications of the syllabus.	⓪	①	②	③	④
	2. Overall, the book has a nice feel.	⓪	①	②	③	④
NA						
☐	B. Methodology					
	3. The activities can be exploited fully.	⓪	①	②	③	④
	4. The acctivities can work well in most classroom situations.	⓪	①	②	③	④
NA						
☐	C. Suitability to learners					
	5. It is compatible with the background knowledge and level of students.	⓪	①	②	③	④
	6. It addresses learning targets.	⓪	①	②	③	④
NA						
☐	D. Physical and utilitarian attributes					
	7. It is appropriately priced.	⓪	①	②	③	④
	8. Its layout is attractive.					
	9. It indicates efficient use of text and visuals.	⓪	①	②	③	④
NA						
☐	E. Efficient layout of supplementary materials					
	10. The book is supported by suitable materials like a workbook, audio, or multimedia.	⓪	①	②	③	④

	11. The book is supported by other materials like review and test units.	⓪	①	②	③	④
	12. There is a useful teacher's guide to aid the teacher.	⓪	①	②	③	④
NA	**II. Learning-teaching content**					
☐	F. General content					
	13. Tasks move from simple to complex.	⓪	①	②	③	④
	14. Tasks are varied.	⓪	①	②	③	④
	15. Tasks support teaching objectives.	⓪	①	②	③	④
	16. The language in the textbook is natural and real.	⓪	①	②	③	④
	17. The material is fairly recent.	⓪	①	②	③	④
NA						
☐	G. Listening					
	18. The book has appropriate listening tasks with well-defined goals.	⓪	①	②	③	④
	19. Tasks are authentic or close to real language situations.	⓪	①	②	③	④
	20. Various listening contexts such as formal vs. informal contexts are considered.	⓪	①	②	③	④
NA						
☐	H. Speaking					
	21. Activities are developed to initiate meaningful communication.	⓪	①	②	③	④
	22. Individual, pair and group work are given equal emphasis.	⓪	①	②	③	④
NA						
☐	I. Reading					
	23. Length is appropriate.	⓪	①	②	③	④
	24. Difficulty level is appropriate.	⓪	①	②	③	④
	25. Texts are interesting.	⓪	①	②	③	④
NA						
☐	J. Writing					
	26. Tasks have achievable goals and take into consideration learner capabilities.	⓪	①	②	③	④
	27. Tasks are interesting.	⓪	①	②	③	④
NA						
☐	K. Vocabulary					
	28. The load (number of new words in each lesson) is appropriate to the level of students.	⓪	①	②	③	④

		⓪	①	②	③	④
	29. There is a good distribution (simple to complex) of vocabulary load across chapters and the whole book.	⓪	①	②	③	④
	30. New words are sufficiently repeated and recycled across the book.	⓪	①	②	③	④
	31. Words are contextualized.	⓪	①	②	③	④
NA ☐	L. Grammar					
	32. Grammar is contextualized.	⓪	①	②	③	④
	33. Grammar items are repeated throughout the book.	⓪	①	②	③	④
NA ☐	M. Pronunciation					
	34. Tasks are useful.	⓪	①	②	③	④
	35. Tasks are interesting.					
NA ☐	N. Exercises					
	36. They have clear instructions.	⓪	①	②	③	④
	37. They are adequate.	⓪	①	②	③	④
	38. They are interesting.	⓪	①	②	③	④
	39. They help students with mixed abilities.	⓪	①	②	③	④

Scores Interpretation Guide

Level	Range	Interpretation
0	0.00-0.80	Negligible usefulness
1	0.81-1.60	Low usefulness
2	1.61-2.80	Moderate usefulness
3	2.81-3.60	High usefulness
4	3.61-4.00	Very high usefulness