



Usefulness of the English Language Teaching Textbook Evaluation Checklist

Vahid Nimehchisalem[#] and Jayakaran Mukundan^{*}

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

ABSTRACT

Textbook evaluation checklists are instruments that help teachers or programme developers evaluate teaching materials before or after using them. This paper presents one of the phases of a project that involved the survey of a group of evaluators' (n=82) views on the usefulness of a newly developed checklist. The questionnaire used was a modified version of an instrument developed to evaluate the usefulness of a writing scale (Nimehchisalem, 2010). Based on the results, the checklist indicates high to very high levels of usefulness. The findings are expected to be very helpful for researchers who may intend to test similar instruments.

Keywords: Textbook evaluation checklist, instrument development, checklist evaluation

INTRODUCTION

Material evaluation checklists are useful instruments that can help teachers who wish to evaluate teaching-learning materials before or after using them. Material evaluation often takes place impressionistically based on evaluators'

intuitive judgment. Checklists, however, help evaluators who lack the experience for an impressionistic evaluation of the material. They are also helpful when evaluators need to work in teams to evaluate materials. The explicit criteria in a checklist enable the evaluator to come up with more objective judgments on the suitability of the material.

Teachers or programme developers often need to evaluate textbooks for the two major reasons of selection and adaptation (Sheldon, 1988). At times, textbooks are evaluated for selection purposes. To choose the most appropriate

ARTICLE INFO

Article history:

Received: 15 October 2012

Accepted: 13 September 2012

E-mail addresses:

nimechie22@yahoo.com (Vahid Nimehchisalem),

jayakaranmukundan@yahoo.com (Jayakaran Mukundan)

^{*} Corresponding author

[#] Author's current affiliation

Department of English, Faculty of Languages and Linguistics,
Universiti of Malaya, 50603 Kuala Lumpur, Malaysia

book for a language programme, one needs to consider a number of books with the needs and specific features of the target students in mind. A textbook may also be evaluated to shed light on its areas of weakness. Once they have been detected, these areas can be improved to increase the effectiveness of the textbook.

One may evaluate textbooks implicitly or explicitly. Implicit evaluation is a common practice among material evaluators. In this type of evaluation, a teacher takes a quick look at the material and impressionistically decides on its suitability. It may be acceptable to evaluate textbooks impressionistically when the teacher has enough experience, but it may be hard to rely on novice teachers' judgements about the suitability of a textbook for a given language programme. In such situations, checklists are used. The explicit criteria in these instruments aid the teachers to carry out a thorough evaluation of the textbook. In addition, checklists are also useful when a panel of experts are involved in evaluating a textbook. It may be hard to reach an agreement in the team evaluation of textbooks. Using checklists can aid the evaluators in such cases to standardize their evaluation and make it more systematic.

Despite their importance, most ELT material evaluation checklists are developed with no proof on their validity or reliability. Some of the checklists commonly used in material evaluation have not been tested for practicality. They are very long that it takes a long time to evaluate language teaching

materials using them. They have items that make it very hard to evaluate materials manually. For example, Skierso's (1991) checklist is a well-established instrument, but it is too long and uneconomical. Additionally, the Skierso checklist also has items that are very hard to rate. For example, there are a few items on the distribution and recycling of the new vocabulary items. It is not practical to analyze such patterns without using software like Wordsmith (Scott, 2004) or Retrotex-E (Mukundan, 2011). In order to provide language teachers and researchers with a valid, reliable and practical instrument to evaluate textbooks, the ELT Textbook Evaluation Checklist was developed using both qualitative and quantitative methods. The next section briefly reviews the development procedure of this instrument.

ELT Textbook Evaluation Checklist

The checklist was first designed based on the review of the available instruments for ELT material evaluation (Mukundan & Ahour, 2010; Mukundan, Hajimohammadi, & Nimehchisalem, 2011). Qualitative and quantitative methods were used to refine the checklist. The qualitative phase comprised a focus group study in which the clarity and inclusiveness of the instrument were refined (Mukundan, Nimehchisalem, & Hajimohammadi, 2011). In the quantitative phase, ELT experts' views were surveyed on the weightage of each domain, and some items were removed based on the results of factor analysis (Mukundan & Nimehchisalem, 2012). Next, the checklist

was operationalized. A five-point Likert style scale was assigned and an interpretation guide was added to it. Appendix A shows the checklist.

Evaluation of Textbook Evaluation Checklists

Textbooks have long been an inseparable part of most language classrooms. It is imperative to evaluate the suitability of these books before, while or after their use. However, it is equally important to evaluate the instruments that are used for evaluating the textbooks. Checklists should be rigorously tested before they are presented as well-established references that can be trusted. Scholars warn about a number of pitfalls that should be avoided in developing their instruments. Tomlinson (2003), for example, warns against the following common mistakes in developing textbook evaluation checklists:

1. mixing evaluation and analysis questions,
2. including two or multiple questions in a single item,
3. developing long, vague, and unanswerable items,
4. asking dogmatic questions, and
5. using items that may be interpreted in a different way by different evaluators.

A review of the available checklists in the literature shows that checklist developers commonly make these mistakes. Mukundan and Ahour (2010) mention examples of checklists which are defective in reference

to the five points mentioned here. They mention an example of multiple questions from Garinger (2002), 'Is the textbook part of a series, and if so would using the entire series be appropriate?' Such items should be avoided in checklists since they will confuse the evaluator and affect the validity and reliability of the evaluation results.

In order to test the usefulness of our checklist, we needed a more systematic framework. For this purpose, we found Bachman and Palmer's (1996) notion of test usefulness suitable. The checklist was, therefore, evaluated using an adaptation of the concept of test usefulness by Bachman and Palmer (1996), to be discussed in the next section.

Usefulness

The concept of usefulness is adapted from Bachman and Palmer (1996). They point out reliability, construct validity, authenticity, interactiveness, impact and practicality as the complementary qualities of any useful test. Out of these, authenticity and interactiveness are more relevant to test tasks. For this reason, only four of the aforementioned qualities in this study were considered, each of which is discussed briefly in this section.

Reliability means consistency of scores, or "scoring validity" (Weir, 2005, p. 22). When a checklist is reliable, it helps different raters to assign almost the same scores for the same textbook. When a checklist helps different evaluators rate the same textbook similarly, it has high inter-rater reliability. In addition, if the items of a checklist indicate

a high degree of consistency, it has internal reliability. Within the scope of this study, reliability means the consistency among the domains of the developed checklist. When developing instruments, reliability can be maximized if the items are clearly stated, and if the ambiguous items are reworded or deleted. Additionally, reliability often rises as the number of its items and/or dimensions increase (Brown & Bailey, 1984; Weir, 2005). Another way to increase reliability is by getting more two or more evaluators to rate the same textbook and considering the average score in the final decision.

Validity means measuring what one claims to be measuring (Cronbach, 1971). Validity may be judgemental or empirical (Richards & Schmidt, 2002). Judgemental validity relies on theory rather than observation. It includes face, content (also context; Weir, 2005), and construct validity. If the checklist appears to measure what it claims to measure, based on the subjective judgement of an observer, it has face validity. If it adequately and sufficiently measures what it sets out to measure, it has content validity. Finally, if it reflects the essential aspects of the theory on which it is based, it has construct validity. Empirical validity, on the other hand, relies on comparing the test with one or more criterion measures and includes criterion-related and consequential validity. If a checklist has criterion-related validity, it indicates high and significant correlations with an established external criterion measure. Finally, if a checklist has consequential validity, the interpretations made by it result in fair and positive social consequences. That is, the users and

stakeholder of a checklist are satisfied with it, it is consequentially valid. In this study, consequential validity is referred to as impact, discussed below.

As the next quality of a useful checklist, impact overlaps with the notion of consequential validity discussed above. A useful checklist should have positive influence on its users and stakeholders. A test can have either micro level or macro level impact on its stakeholders (Bachman & Palmer, 1996). At a micro level, a checklist can influence teachers who can use it to diagnose problematic areas in their teaching material and make adjustments in their future instructions. If a checklist is used in this way, it can have desirable impact (or positive washback effect) on the instruction. On the other hand, a useful can have a positive effect at a macro level by making programme developers aware of the suitability of the materials they have developed. According to this useful information, they can adjust the syllabi to the target learners' needs.

Bachman and Palmer (1996) define practicality as "the relationship between the resources that will be required in the design, development and use of the test and the resources that will be available for these activities" (p. 36). These resources include: (1) teachers and administrators (or, human resources), (2) equipment and materials, and (3) time (Bachman & Palmer, 1996). Instrument developers should remember that the issue of practicality overrides all the previously mentioned qualities. That is, they should not be so obsessed with the reliability and validity of their instruments that they

forget how it will work in real testing situations. If the prospective stakeholders' limitations and resources are not considered, they may refuse to use the checklist even if it is the most valid and reliable instrument available and will prefer a second best checklist that they find practical.

RESEARCH OBJECTIVE

In the light of the four test qualities discussed above, this paper aimed at investigating a number of English language teachers' views on the usefulness of the developed checklist. The teachers were surveyed on the reliability, relevance, impact, and practicality of the checklist. The following research questions were postulated to address this objective.

RESEARCH QUESTIONS

The following research questions were posed to address the aforementioned objective:

1. To what extent do the respondents regard the ELT Textbook Evaluation Checklist as a reliable instrument?
2. To what extent do the respondents regard the ELT Textbook Evaluation Checklist as a valid instrument?
3. To what extent do the respondents regard the English Language Teaching (ELT) Textbook Evaluation Checklist useful in reference to its impact and practicality?
4. How can the English Language Teaching (ELT) Textbook Evaluation Checklist be improved?

In order to provide answers to the first three questions, the survey method was used whereas the qualitative method was used to address the final research question.

MATERIALS AND METHOD

Previous attempts were made to improve the reliability and validity of the checklist through qualitative (Mukundan *et al.*, 2011; Mukundan *et al.*, 2011) and quantitative (Mukundan & Nimehchisalem, 2012) methods. This research was an attempt to explore the usefulness of the checklist with the help of a questionnaire. The survey method was used to conduct this study.

Respondents

The respondents comprised of 82 male and female Vietnamese English language teachers with a bachelor degree in Teaching English as a Second Language. They had a minimum of two years of teaching experience. The respondents were given the questionnaire right after they had used the checklist to evaluate two textbooks.

Instrument

The instrument that was used for collecting data was a modified version of a questionnaire called the Argumentative writing Scale Evaluation Questionnaire (Nimehchisalem, 2010). The modified version of the instrument is called the 'ELT Textbook Evaluation Checklist Evaluation Questionnaire'. The only demographic information elicited from the respondents was their years of teaching experience

to make the instrument as economical as possible. The questionnaire is a five-point Likert style instrument, where a value of 1 signifies ‘strongly disagree’; 2 ‘disagree’; 3 ‘not sure’; 4 ‘agree’; and 5 ‘strongly agree’. The questionnaire consists of 13 items, followed by a final open-ended question. Appendix B presents the instrument of the study.

The questionnaire was primarily developed based on Bachman and Palmer’s (1996, pp. 19-40) definition of test usefulness, which includes reliability, construct validity, impact and practicality. Slight modifications were made to the original questionnaire to make it fit to

evaluate the usefulness of a textbook evaluation checklist. The modifications included changing words like ‘writing scale’ and ‘students’ written works’ to ‘checklist’ and ‘textbook’ respectively in order to make it relevant for checklist evaluation. In addition, three reverse items were added to the questionnaire to ensure the honesty of the respondents. Table 1 summarizes each item in the questionnaire, along with the dimension of usefulness that it addresses.

As the table shows, items 1-4, and 13 are related to the impact and practicality of the checklist. Meanwhile, reliability is addressed by items 6-11, whereas items 5, 6, 9, and 12 deal with validity. Overall,

TABLE 1
Components of test usefulness covered in the questionnaire

Item	Component
1 I found it easy to work with the checklist.	Impact, practicality
2 It is tiring to evaluate textbooks using the checklist.	Impact, practicality
3 I will use this checklist to evaluate the textbooks that I use in my own classes.	Impact, practicality
4 I recommend using this checklist to my colleagues.	Impact, practicality
5 The checklist fully covers the aspects of textbook evaluation construct.	Validity
6 I find my personal judgment of the textbook in line with the score assigned by the checklist.	Reliability, Validity
7 The checklist helped me draw a clear line between the books that seemed to be of different qualities.	Reliability
8 All the terms in the checklist are clear and easy to understand.	Reliability
9 Some items are vague.	Reliability, Validity
10 Overall, I find the checklist a reliable instrument.	Reliability
11 The scoring guide is clear.	Reliability
12 There are certain important dimensions in textbook evaluation that are missing in the checklist.	Validity
13 Overall, I am satisfied with this checklist.	Impact, practicality
14 What changes do you think can be made on the checklist to improve it?	All

a positive response to all the items in the questionnaire would imply the respondents' satisfaction of the checklist, excluding items 2, 9, and 12 that are reverse items. Based on Cronbach's Alpha test of reliability, an acceptable coefficient .76 was recorded for the instrument.

Research Procedure

The respondents were first asked to evaluate two English language textbooks using the ELT Textbook Evaluation Checklist. After they had evaluated the textbooks, they rated the checklist itself using the English Language Teaching Textbook Evaluation Checklist Evaluation Questionnaire (Appendix B).

Data Analysis

The questionnaires were collected. The reverse items were checked to see if the respondents had been honest in answering the questions. The questionnaires in which the reverse items had been marked similar to the other items were disregarded. As a result, 82 questionnaires were selected to be coded for data entry. SPSS Version 16 was used to analyze the data. Descriptive statistical procedures, which included frequency, means, and standard deviation, were followed. The values of the reverse items were transformed; for example, an item marked as '5' was recoded as '1' and vice versa before analysis.

The scores of each domain (impact and practicality, items 1-4, 13; validity, items 5, 6, 9, and 12; reliability, items 6-11) were added, and their percentages were computed.

The percentages were then categorized into five levels based on Guilford's (1950) Rule of Thumb to facilitate the interpretation of the respondents' ratings. Following this rule, 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' levels of the teachers' satisfaction with the checklist.

RESULTS

This study aimed at surveying a group of teachers' views on the usefulness of a recently developed checklist. For this purpose, a number of research questions were posed. The research questions covered the components of checklist usefulness. This section presents and discusses the results of the survey with a focus on each question subsequently. The results have been presented as means (*M*) and standard deviations (*SD*).

Impact and Practicality

The first research question concerned the impact and practicality of the checklist. These qualities were evaluated through five items including, 'I found it easy to work with the checklist' (item 1); 'It is tiring to evaluate textbooks using the checklist' (item 2); 'I will use this checklist to evaluate the textbooks that I use in my own classes' (item 3); 'I recommend using this checklist to my colleagues' (item 4); and 'Overall, I am satisfied with this checklist' (item 13). Table 2 summarizes the descriptive statistics results of the teachers' responses to the aforementioned items:

TABLE 2
Descriptive statistics results of impact and practicality items

	Item1	Item2 (reversed)	Item3	Item4	Item13
Mean (upon 5)	3.9390	3.8659	3.9756	4.0000	4.2561
Std. Deviation	.89370	.73303	.71966	.73703	.62482
strongly disagree (%)	2.4	00	00	00	00
disagree (%)	6.1	00	1.2	4.9	1.2
unsure (%)	9.8	34.1	23.2	12.2	6.1
agree (%)	58.5	45.1	52.4	61.0	58.5
strongly agree (%)	23.2	20.7	23.2	22.0	34.1

Items 1, 2, 3, 4 and 13 (in table 2) were rated at $M=3.9$, $SD=.89$; $M=3.9$, $SD=.73$; $M=4$, $SD=.72$; $M=4$, $SD=.74$; and $M=4.3$, $SD=.62$, respectively. According to these results, a majority of the teachers (on average about 55%) ‘agreed’ on the practicality and impact of the checklist. The item that obtained the lowest level of satisfaction was the second item. Regarding this item, one in three teachers felt ‘unsure’ whether using the checklist was tiring or not. The rest (65.8%) believed that it was not tiring to work with the instrument. The most promising feedback was achieved from item 13, which dealt with the respondents’ overall satisfaction of the instrument. As shown by the results, over 92% of the teachers ‘agreed’ or ‘strongly agreed’ that overall they were satisfied with the checklist.

Based on Guilford’s (1950) Rule of Thumb, the teachers’ ratings were transformed into the five satisfaction categories of ‘Negligible’, ‘Low’, ‘Moderate’, ‘High’, and ‘Very high’. Table 3 presents the teachers’ satisfaction with the impacts and practicality of the checklist:

As it can be observed, about three in four teachers (73.2%) were ‘highly’ satisfied with the developed instrument. The rest were either ‘moderately’ (9.8%) or ‘very highly’ (17.1%) satisfied with it. These findings are promising and provide proof for high consequential validity of the ELT Textbook Evaluation Checklist.

TABLE 3
Teachers’ satisfaction levels in reference to impact and practicality of the checklist

	Frequency	Percent	Cumulative Percent
Moderate	8	9.8	9.8
High	60	73.2	82.9
Very high	14	17.1	100.0

Validity

The second research question dealt with the validity of the checklist. The items focusing on the validity of the checklist comprised, ‘The checklist fully covers the aspects of textbook evaluation construct’ (item 5); ‘I find my personal judgment of the textbook in line with the score assigned by the checklist’ (item 6); ‘Some items are vague’ (item 9);

and ‘There are certain important dimensions in textbook evaluation that are missing in the checklist’ (item 12). As mentioned before, items 9 and 12 were reverse items and therefore had to be reversed before the analysis. Table 4 presents the means and standard deviations of the ratings assigned for these items.

As revealed by the results in table 4, items 5, 6, 9, and 12 respectively achieved means and standard deviations of $M=4.1$, $SD=.64$; $M=3.9$, $SD=.67$; $M=3.8$, $SD=.73$; and $M=3.7$, $SD=.71$. Overall, majority of the teachers (on average, 54.9%) ‘agreed’ on the validity of the instrument. The item that achieved the highest percentage concerned the inclusiveness and construct validity of the instrument, that is, it fully covers the construct of textbook evaluation.

Additionally, as in the case of the previous research question, the results were transformed into five satisfaction categories based on Guilford’s (1950) Rule of Thumb. Table 5 presents the teachers’ satisfaction with the checklist in relation to its validity.

As it can be seen in Table 5, a majority of the teachers (73.2%) were ‘highly’ satisfied with the validity of the checklist.

A few (13.4%) were ‘moderately’ satisfied while some (13.4%) were ‘very highly’ satisfied with its validity. However, none of the teachers indicated low levels of satisfaction in relation to the validity of the checklist. These results are in line with the findings of the previous research question (that concerned the practicality and impacts of the checklist). Exactly the same percentage of the respondents (73,2%) rated both domains as highly satisfactory, while the remaining proportion indicated either moderate or very high levels of satisfaction. The findings, therefore, provide further proof for the validity of the ELT Textbook Evaluation Checklist.

TABLE 5
Teachers’ satisfaction levels in reference to validity of the checklist

	Frequency	Percent	Cumulative Percent
Moderate	11	13.4	13.4
High	60	73.2	86.6
Very high	11	13.4	100.0

TABLE 4
Descriptive statistics results of validity items

	Item5	Item6	Item9 (reversed)	Item12 (reversed)
Mean (upon 5)	4.1341	3.8902	3.7927	3.7195
Std. Deviation	.64334	.66678	.73262	.70753
strongly disagree (%)	00	00	00	00
disagree (%)	1.2	1.2	4.9	4.9
unsure (%)	11.0	24.4	24.4	42.7
agree (%)	61.0	58.5	57.3	42.7
strongly agree (%)	26.8	15.9	13.4	14.6

Reliability

The third research question focused on the reliability of the checklist. The items investigating the teachers' views towards the reliability of the instrument were, 'I find my personal judgment of the textbook in line with the score assigned by the checklist' (item 6); 'The checklist helped me draw a clear line between the books that seemed to be of different qualities' (item 7); 'All the terms in the checklist are clear and easy to understand' (item 8); 'Some items are vague' (item 9); 'Overall, I find the checklist a reliable instrument' (item 10); and 'The scoring guide is clear' (item 11). As mentioned previously, item 9 was a reverse item and therefore reversed before the analysis. Table 6 presents the means and standard deviations of the ratings that the respondents assigned for these items.

According to the results in table 6, items 6 to 11 achieved means and standard deviations of $M=3.9, SD=.67$; $M=4.1, SD=.63$; $M=3.9, SD=.89$; $M=3.8, SD=.73$; $M=4.2, SD=.61$; and $M=3.8, SD=.67$, respectively. Like practicality and validity, most of the respondents (on average,

54.7%) 'agreed' that the instrument was reliable. The item that achieved the highest percentage was item 7 which focused on the clarity of the checklist items. About one in four respondents believed that the checklist aided them to draw a line between the suitable and unsuitable textbooks.

Moreover, transformation of the results into five satisfaction categories demonstrated the respondents' level of satisfaction with the checklist in reference to its reliability. Table 7 summarizes these results.

TABLE 7
Teachers' satisfaction levels in reference to reliability of the checklist

	Frequency	Percent	Cumulative Percent
Moderate	9	11.0	11.0
High	61	74.4	85.4
Very high	12	14.6	100.0

As depicted in Table 7, about three in four teachers (74.4%) were 'highly' satisfied with the reliability of the checklist. The remaining few were either 'moderately' (11%) or 'very highly' (14.6%) satisfied with the reliability of the checklist. This

TABLE 6
Descriptive statistics results of reliability items

	Item6	Item7	Item8	Item9 (reversed)	Item10	Item11
Mean (upon 5)	3.8902	4.1037	3.9146	3.7927	4.2195	3.8049
Std. Deviation	.66678	.63215	.89168	.73262	.60908	.67475
strongly disagree (%)	00	00	00	00	00	00
disagree (%)	1.2	1.2	4.9	4.9	00	1.2
unsure (%)	24.4	11.0	29.3	24.4	9.8	30.5
agree (%)	58.5	63.4	35.4	57.3	58.5	54.9
strongly agree (%)	15.9	24.4	30.5	13.4	31.7	13.4

meant that none of the teachers indicated low levels of satisfaction in relation to the reliability of the instrument. These results confirm the findings of both previous research questions, providing proof for high reliability of the ELT Textbook Evaluation Checklist. Table 8 presents a summary of the results for each component of checklist usefulness.

Table 8 illustrates that the means reported for each component were equal to or more than 4. This suggests that, on average, the teachers 'agreed' that the checklist has 'impact and practicality', 'validity', and 'reliability'. Furthermore, the respondents' overall satisfaction of the checklist in reference to its usefulness is presented in Table 9.

Overall, most of the teachers (78%) considered the checklist as a 'highly' useful instrument for evaluating the suitability of English language teaching textbooks. Some (12.2%) indicated that its usefulness was 'very high', while a minority (9.8%) regarded its usefulness as just 'moderate'.

As it is evident from the findings presented so far, the teachers had a positive attitude towards the checklist, but still there were a few of them who regarded it to be moderately useful. The open-ended question

at the end of the questionnaire gave a chance for these respondents to provide useful feedback for the developers to further refine the checklist. The next section presents the results of this open-ended question.

TABLE 9
Teachers' perceptions towards the usefulness of the checklist

	Frequency	Percent	Cumulative Percent
Moderate	8	9.8	9.8
High	64	78.0	87.8
Very high	10	12.2	100.0

Qualitative Findings

This section presents the findings of the final research question. The teachers were asked whether they had any recommendations for further improvement of the checklist. As it was expected, only a few of the respondents (about 10%) responded to the open-ended question at the end of the questionnaire.

There was a comment on the order of the evaluative criteria in the checklist. One of the teachers recommended moving the items on language skills to the beginning of the checklist. The order in which the domains of an instrument appear can indicate their importance. This teacher believes that

TABLE 8
Descriptive statistics results for the checklist usefulness (n=82)

	Impact and Practicality categories	Validity categories	Reliability categories	Total
Mean (upon 5)	4.0732	4.0000	4.0366	4.0244
Std. Deviation	.51593	.52116	.50784	.47077
Minimum	3.00	3.00	3.00	3.00
Maximum	5.00	5.00	5.00	5.00

language skills are very important and that they should be emphasized over other criteria. Our previous findings, however, showed that almost all the evaluative criteria in the checklist have an equal level of importance (Mukundan & Nimehchisalem, 2012). This comment was, therefore, overlooked.

There were quite a number of comments on adding certain components of language teaching to the checklist. One of the respondents mentioned that the notion of style should be considered in language textbook evaluation. Similarly, another respondent pointed out the formal and informal varieties of language. One of the respondents pointed out the importance of various skills, including problem-solving, note-taking, reporting, summarizing, and the like. In addition, another respondent recommended adding an item on creativity, like 'Do the tasks stimulate students' creativity and imagination?' There was also a comment on adding another item to the listening section which would focus on the presentation of the various English accents in the listening section. A respondent suggested adding a qualitative part to the checklist. Finally, there was a comment on the flexibility of the checklist. As one of the teachers recommended, "The checklist should be more flexible and allow the teachers to skip some sections that may not be relevant to their teaching context". There are textbook evaluation checklists that allow for disregarding certain items that may turn out irrelevant to particular situations. Skierso (1991) did so by adding

a column called 'not applicable' to the front of each item. If the evaluator finds the items irrelevant to his/her present context, he/she may just ignore it.

In addition to the comments for moving or adding more items, there were also recommendations on rewording some of the items. As one of the respondents suggested, items like 'It is compatible to students' background knowledge and level' could be reworded as 'It is compatible to the background knowledge and level of your students'. The respondent explained that this would urge the evaluators to examine the textbook in reference to their present setting. In this way, the checklist would encourage the evaluator to focus on her/his own learners. Another respondent considered the first item in the checklist; that is, 'The textbook matched the specifications of the syllabus' too broad and, therefore, a bit vague. Two of the respondents assumed that the items had to be worded more simply to make it easier to comprehend them.

DISCUSSION

Based on the results of the quantitative data analysis, it was concluded that the respondents regarded the checklist as a useful instrument. The findings thus provide some evidence on the impact and practicality, validity, as well as reliability of the checklist. However, before the checklist could be used confidently by researchers and language teachers as a valid and reliable instrument further research seemed necessary on a panel of experts evaluation of its validity and practicality.

The recommendations made by some of the respondents in response to the final research question would help the researchers make the checklist more comprehensive and precise. This would increase the validity and reliability of the instrument. The idea of making the checklist more flexible, for example, sounded very important. Indeed, the researchers considered adding a line to the instructions to allow the evaluator to ignore some of the items. This would, however, make the checklist complicated like Skierso's checklist. When an instrument is too complicated, it is not practical. What the present researchers assumed more important, therefore, was the practicality and economy of the checklist. As mentioned before, there was a suggestion for adding a qualitative part to the checklist. Open-ended questions would definitely result in richer and more in-depth evaluations of textbooks. Novice evaluators would, however, find it hard to work with them as it is quite difficult to interpret the responses to qualitative checklists. Such comments were, therefore, ignored since it would affect the practicality of the checklist.

There were recommendations for adding skills like problem-solving, note-taking, reporting, and the like. This comment alone would add a minimum of four items to the

checklist. Not to mention that adding such evaluative criteria to the checklist would result in unfair evaluations of elementary level textbooks that rarely emphasize such skills. This argument sounds true for the comment on adding an item on the presentation of the various English accents. Such an item sounds more appropriate for advanced level textbooks. Indeed, in particular occasions, certain English accents may be preferred to others considering the learners' needs. It would be, therefore, unfair to rate a textbook that presents various accents higher than one that does not.

Apart from the issue of practicality, adding any new item to the instrument would lengthen the checklist and reduce its economy. The longer the checklist, the more costly its use and administration will be. Also, it usually takes a considerable amount of time to evaluate textbooks using long checklists. Mukundan and Ahour (2010), in their extensive review of textbook evaluation checklists, compared the number of tokens (that is, the total number of words) in each checklist. Table 10 shows the number of running words in the three shortest and the three longest checklists reviewed by Mukundan and Ahour (2010).

The number of running words in our checklist was 361; that is, about 3 times

TABLE 10
Running words in some textbook evaluation checklists

Shortest checklists		Longest checklists	
Checklist	Running words	Checklist	Running words
Tucker (1978)	113	Rivers (1981)	1981
Ur (1996)	126	Litz (2005)	2534
Byrd <i>et al.</i> (2001)	163	Skierso (1991)	4553

longer than the shortest (i.e., Tucker, 1978), and yet, more than 12 times shorter than the longest checklist (that is, Skierso, 1991). The ELT Textbook Evaluation Checklist, therefore, can be regarded as one of the most economical checklists.

CONCLUSION

The objective of this research was to survey a group of English language teachers' views on the usefulness of a newly developed textbook evaluation instrument called the ELT Textbook Evaluation Checklist. As indicated by the results, the respondents generally agreed that the checklist is a useful instrument. Furthermore, the respondents also made comments on how the checklist could be improved. Some of these comments were implemented in the checklist. It is expected that the instrument can help language teachers and researchers in evaluating the suitability of English language teaching textbooks.

In this study, a questionnaire was used to evaluate the usefulness of the developed checklist. Researchers and instrument developers can adapt this questionnaire to elicit experts' feedback on similar instruments. The questionnaire is based on Bachman's and Palmers' concept of test usefulness, and therefore, has a sound theoretical foundation. The unique feature of this questionnaire is its economy. It only has 13 items. Additionally, its reverse items help the researcher eliminate the questionnaires that have not been honestly attended to. A final merit of this instrument is the open-

ended question at the end which helps the researcher to find out an in-depth account of what some evaluators' real judgment.

One of the main limitations of this study is that it only considers the views and perceptions of a group of English language teachers. Another study that focuses on a panel of experts' evaluation of the checklist can provide more precise and reliable account of its usefulness. Further research is also required to test the concurrent validity of the checklist in reference to well-established instruments in the area. Furthermore, the economy of the instrument can be compared with that of other similar instruments by having the same teachers use different checklists for evaluating the same textbook.

REFERENCES

- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press.
- Brown, J. D., & Bailey, K. M. (1984). A categorical instrument for scoring second language writing skills. *Language Learning*, 34(1), 21-42.
- 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). Boston: Heinle & Heinle.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational Measurement (2nd Ed.)*. Washington DC: American Council on Education.
- Garinger, D. (2002). Textbook selection for the ESL classroom. *Center for Applied Linguistics Digest*. Retrieved June, 1st, 2012, from <http://www.cal>

- org/resources/digest/0210garinger.html.
- Guilford, J. P. (1950). *Fundamental statistics in psychology and education* (2nd Ed.). New York: McGraw-Hill.
- Litz, D. R. A. (2005). Textbook evaluation and ELT management: A South Korean case study. *Asian EFL Journal* (Thesis section). Retrieved June 2nd, 2012, from <http://www.asian-efl-journal.com/thesis.php>.
- Mukundan, J. (2009). *ESL textbook evaluation. A composite framework*. Köln, Germany: Lambert Academic Publishing.
- Mukundan, J. (2011). Retrotext-E 1.0: The Beginnings of Computer-based ELT Textbook Evaluation. In J. Mukundan & V. Nimehchisalem (Eds.), *Reading on ELT Materials V* (pp. 270-280). Serdang: Faculty of Educational Studies, Universiti Putra Malaysia.
- 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. (2011). Developing an English language textbook evaluation checklist. *Contemporary Issues in Education Research*, 4(6), 21-27.
- Mukundan, J., & Nimehchisalem, V. (2012). Evaluative criteria of an English language textbook evaluation checklist. *Journal of Language Teaching Research*, 3(6), 1128-1134.
- Mukundan, J., Nimehchisalem, V., & Hajimohammadi, R. (2011). Developing an English language textbook evaluation checklist: A focus group study. *International Journal of Humanities and Social Science*, 1(12), 100-105.
- Nimehchisalem, V. (2010). *Developing an analytic scale for argumentative writing of students in a Malaysian public university* (Unpublished doctoral dissertation). Universiti Putra Malaysia, Serdang, Malaysia.
- Richards, J. C., & Schmidt, R. (2002). *Longman Dictionary of Language Teaching & Applied Linguistics*. London: Pearson Education.
- Rivers, W. M. (1981). *Teaching foreign language skills*. Chicago: University of Chicago Press.
- Scott, M. (2004). WordSmith Tool Manual Version 3.0., Retrieved April 20th, 2011, from <http://www.lexically.net/wordsmith/version3/index.htm>.
- 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.
- Tomlinson, B. (2003). Materials evaluation, in B. Tomlinson (Ed.), *Developing materials for language teaching* (pp. 15-36). London: Continuum.
- Tucker, C. A. (1978). Evaluating beginning textbooks. In H. S. Madsen, & J. D. Brown, (Eds.), *Adaptation in language teaching* (pp. 219-37). Rowley, Mass: Newbury.
- Ur, P. (1996). *A course in language teaching: Practice and theory*. Cambridge: Cambridge University Press.
- Weir, C. J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Hampshire, UK: Palgrave MacMillan.

APPENDIX A

ENGLISH LANGUAGE TEACHING TEXTBOOK EVALUATION CHECKLIST

Evaluator:

Teaching experience: (years)

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 in relation to syllabus and curriculum	
1. It matches to the specifications of the syllabus.	① ① ② ③ ④
B. Methodology	
2. The activities can be exploited fully and can embrace the various methodologies in ELT.	① ① ② ③ ④
3. Activities can work well with methodologies in ELT.	① ① ② ③ ④
C. Suitability to learners	
4. It is compatible to background knowledge and level of students.	① ① ② ③ ④
5. It is compatible to the socio-economic context.	① ① ② ③ ④
6. It is culturally accessible to the learners.	① ① ② ③ ④
7. It is compatible to the needs of the learners.	① ① ② ③ ④
D. Physical and utilitarian attributes	
8. Its layout is attractive.	① ① ② ③ ④
9. It indicates efficient use of text and visuals.	① ① ② ③ ④
10. It is durable.	① ① ② ③ ④
11. It is cost-effective.	① ① ② ③ ④
12. Its size is appropriate.	① ① ② ③ ④
13. The printing quality is high.	① ① ② ③ ④

Usefulness of an English Textbook Evaluation Checklist

E. Efficient outlay of supplementary materials					
14. The book is supported efficiently by essentials like audio-materials.	①	①	②	③	④
15. There is a teacher's guide to aid the teacher.	①	①	②	③	④
II. Learning-teaching content					
A. General					
16. Tasks move from simple to complex.	①	①	②	③	④
17. Task objectives are achievable.	①	①	②	③	④
18. Cultural sensitivities have been considered.	①	①	②	③	④
19. The language in the textbook is natural and real.	①	①	②	③	④
20. The situations created in the dialogues sound natural and real.	①	①	②	③	④
21. The material is up-to-date.	①	①	②	③	④
22. It covers a variety of topics from different fields.	①	①	②	③	④
23. The book contains fun elements.	①	①	②	③	④
B. Listening					
24. The book has appropriate listening tasks with well-defined goals.	①	①	②	③	④
25. Instructions are clear.	①	①	②	③	④
26. Tasks are efficiently graded according to complexity.	①	①	②	③	④
27. Tasks are authentic or close to real language situations.	①	①	②	③	④
C. Speaking					
28. Activities are developed to initiate meaningful communication.	①	①	②	③	④
29. Activities are balanced between individual response, pair work and group work.	①	①	②	③	④
30. Activities motivate students to talk.	①	①	②	③	④
D. Reading					
31. Texts are graded.	①	①	②	③	④
32. Length is appropriate.	①	①	②	③	④
33. Texts are interesting.	①	①	②	③	④
E. Writing					
34. Tasks have achievable goals and take into consideration learner capabilities.	①	①	②	③	④
35. Models are provided for different genres.	①	①	②	③	④
36. Tasks are interesting.	①	①	②	③	④

<p>F. Vocabulary</p> <p>37. The load (number of new words in each lesson) is appropriate to the level.</p> <p>38. There is a good distribution (simple to complex) of vocabulary load across chapters and the whole book.</p> <p>39. Words are efficiently repeated and recycled across the book.</p> <p>40. Words are contextualized.</p>	<p>① ① ② ③ ④</p> <p>① ① ② ③ ④</p> <p>① ① ② ③ ④</p> <p>① ① ② ③ ④</p>
<p>G. Grammar</p> <p>41. The spread of grammar is achievable.</p> <p>42. The grammar is contextualized.</p> <p>43. Examples are interesting.</p> <p>44. Grammar is introduced explicitly.</p> <p>45. Grammar is reworked implicitly throughout the book.</p>	<p>① ① ② ③ ④</p> <p>① ① ② ③ ④</p> <p>① ① ② ③ ④</p> <p>① ① ② ③ ④</p> <p>① ① ② ③ ④</p>
<p>H. Pronunciation</p> <p>46. It is contextualized.</p> <p>47. It is easy to learn.</p>	<p>① ① ② ③ ④</p> <p>① ① ② ③ ④</p>
<p>I. Exercises</p> <p>48. They have clear instructions.</p> <p>49. They are adequate.</p> <p>50. They help students who are under/over-achievers.</p>	<p>① ① ② ③ ④</p> <p>① ① ② ③ ④</p> <p>① ① ② ③ ④</p>

APPENDIX B**ENGLISH LANGUAGE TEACHING TEXTBOOK EVALUATION CHECKLIST
EVALUATION QUESTIONNAIRE**

This questionnaire has been developed to evaluate the English Language Teaching Textbook Evaluation Checklist based on your judgment of its quality. As an evaluator who used the checklist, you are kindly requested to mark the spaces in front of the statements below that best describe your evaluation of the checklist according to the key provided below:

1. Strongly disagree
2. Disagree
3. Neither agree nor disagree
4. Agree
5. Strongly agree

You are also requested to provide answers for question 14.

Item	1	2	3	4	5	Comments
1. I found it easy to work with the checklist.						
2. It is tiring to evaluate textbooks using the checklist.						
3. I will use this checklist to evaluate the textbooks that I use in my own classes.						
4. I recommend using this checklist to my colleagues.						
5. The checklist fully covers the aspects of textbook evaluation construct.						
6. I find my personal judgment of the textbook in line with the score assigned by the checklist.						
7. The checklist helped me draw a clear line between the books that seemed to be of different qualities.						
8. All the terms in the checklist are clear and easy to understand.						
9. Some items are vague.						
10. Overall, I find the checklist a reliable instrument.						
11. The items have been arranged according to their importance OR have been weighted.						
12. There are certain important dimensions in textbook evaluation that are missing in the checklist.						
13. Overall, I am satisfied with this checklist.						
14. What changes do you think can be made on the checklist to improve it?						
Total: /65						

Thank you very much for your cooperation.

