

## Evaluating the Validity and Economy of the English Language Teaching Textbook Evaluation Checklist

*Jayakaran Mukundan and Vahid Nimehchisalem*

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

---

**Abstract:** This paper presents a part of an on-going project that aims at developing an instrument to evaluate English language teaching textbooks, called the English Language Teaching (ELT) Textbook Evaluation Checklist. The present study sought to evaluate the validity, reliability and economy of the developed checklist. It also aimed to find out to what extent the users of the checklist were satisfied with it. Two ELT experts used the checklist to evaluate the same textbook. The inter-rater reliability results showed acceptable reliability coefficients ( $r=.88$ ). The checklist was tested for the correlation of its results with those of a well-established instrument [1] and a high correlation ( $r=.77$ ) was observed. The instrument also proved to be economical. On average, it took the two experts almost three times shorter as compared to the time they spent using the Skierso checklist. The evaluators turned out to be highly satisfied with the checklist and highly agreed on its usefulness. Despite the promising findings of this study, further research is needed to shed light on the possible problems of the checklist and improve its validity, reliability and practicality.

**Key words:** English language textbook evaluation • Evaluation checklists • Checklist validity

---

### INTRODUCTION

As language teachers, we may have to evaluate textbooks either for selecting them or for adapting them [2]. Textbooks may be evaluated implicitly (following the evaluator's impressionistic judgment) or explicitly (using an instrument). An instrument that is used for evaluating textbooks is referred to as an evaluation checklist, or simply, a checklist. The criteria in a checklist can facilitate the evaluation of textbooks and make it more objective.

Many checklists are available in the literature; however, the problem is that most of them have not been validated. Although there are some checklists which have been tested for their validity and reliability, they lack practicality. While some are too long which reduces their economy, others are too sophisticated which makes it very challenging for novice teachers to use them. Therefore, there is a need for a checklist that is valid, reliable and practical. With this objective, a research

project commenced in 2010, resulted in a tentative checklist in 2011 and then went through a series of field-tests which have improved the quality of the developed checklist. A synopsis of this project is provided in this section before presenting the objective of the present study.

The related literature on ELT textbook evaluation was first reviewed. Mukundan and Ahour [3] present the results of this review which focuses on the checklists developed in the last four decades. Based on this review, a tentative checklist (including 27 items with a 5-point Likert style scale) was developed [4]. Next, a focus group study [5] and a survey [6] were conducted to help the researchers refine the checklist. As a result of these studies, the checklist evolved into a 50-item instrument [6]. A follow-up study was conducted with a focus on the usefulness of the checklist and positive results were obtained [7]. The limitation of that study, however, was that it relied only on a group of English language teachers' perceptions to determine the usefulness of

the checklist. The present study was, therefore, an attempt to test the reliability, validity and economy of the checklist in a more empirical manner. The following research questions were addressed to meet the aforementioned objective:

- How reliable is the ELT Textbook Evaluation Checklist?
- To what extent do the scores assigned using the ELT Textbook Evaluation Checklist correlate with the scores assigned by the Skierso checklist?
- How economical is the ELT Textbook Evaluation Checklist?
- To what extent are the experts who used the checklist satisfied with it?

In the development of any instrument, there are certain issues that should be considered. Among them, reliability, validity and practicality stand out. A reliable instrument has 'scoring validity' [8]. It can aid different evaluators assign similar scores when they are rating the same material (inter-rater reliability). When an instrument is valid, it tests what it claims to be testing [9]. That is, the content of the items in the instrument should be appropriate for what it claims to measure (content validity). The validity of an instrument should be supported both theoretically and empirically [10]. This means that the construct and sub-constructs (or domains) of the instrument should be based on theory (construct validity). It also means that the scores assigned by the instrument should indicate a high correlation with those assigned using other well-established instruments (concurrent validity). It is possible to have instruments that have high reliability and validity but are not very popular among their end users. This is often true when the instrument lacks practicality; that is, it is too long, too complicated, or too costly to be used. A useful checklist, therefore, is a practical instrument that helps the evaluator provide reliable and valid judgment on the suitability of the textbook under examination.

At the heart of any checklist lie the criteria based on which textbooks will be evaluated. These criteria could vary depending on different factors, one of which could be the ELT methods followed in the textbook. That is why one may come across with different criteria when comparing different checklists. Some of the common criteria in most textbook evaluation checklists include novelty, attractiveness, authenticity, communicative purposes, as well as intellectual, aesthetic and emotional involvement [11], strength, lightness and transparency

[12], fitness to the target curriculum, physical characteristics and teachability [13]. It would be inappropriate to evaluate different textbooks based on a global set of criteria due to the variations in learning-teaching contexts. Some universal features have been suggested in the literature, however, including, approach, content presentation, physical make-up and administrative concerns [14]. An extensive review of the criteria used in the checklists is presented by Mukundan and Ahour [3].

## MATERIAL AND METHODS

The research questions were addressed through both qualitative and quantitative methods. This section elaborates on the evaluators, the material and the instruments used for data collection as well as the research procedure.

**Evaluators:** The evaluators who used the checklists to rate the textbooks were two experienced ELT experts. Both held PhD degrees in Teaching English as a Second Language (TESL). They had a minimum of 12 years of teaching experience. The evaluators had been colleagues for the past five years, teaching almost similar materials in the same language center.

**Material:** The textbook that was selected to be evaluated using the checklists was the *New Headway* (2<sup>nd</sup> edition) [15]. Both the evaluators had an experience of using this book for a minimum of five years in the same learning-teaching context.

**Instruments:** Two instruments were used in this study. The first instrument was Skierso's checklist [1]. The instrument has two parts, the first examines the textbook and the second focuses on the teacher's manual. In this study, only the first part was used. It contains six subsections, including 'bibliographical data', 'aims and goals', 'subject matter', 'vocabulary and structures', 'exercises and activities, as well as 'layout and physical makeup'. The instrument has a five-point scale of zero (signifying 'totally lacking') to four (representing 'excellent'). The Skierso checklist was developed based on an extensive review of the related literature and is known as one of the most well-established checklists in the literature [16]. The Skierso checklist had parts that were missing in the ELT Textbook Evaluation Checklist and *vice versa*. Therefore, only the matching items were considered for analysis purposes (Table 1).

Table 1: Overlapping sections in the Skierso and the English Language Teaching (ELT) Textbook Evaluation Checklists

Skierso Checklist (Textbook section)	ELT Textbook Evaluation Checklist
Aims and Goals (B.1-5)	Suitability to learners (C.4-7), syllabus specifications (A.1), & general content (F.17)
B.1 target students' specifications	C.4 compatibility to background knowledge and level of students
B.2 matching to students' needs	C.7 compatibility to the needs of the learners
B.3 matching to syllabus requirements	A.1 matching to the specifications of the syllabus
B.4 compliance with overall educational concerns	C.5 compatibility to the socio-economic context
	C.6 cultural accessibility to the learners
B.5 feasibility	F.17 & L.41 achievable task objectives & grammar
Subject Matter (C)	Items from various sections (F, G, I, & L)
C.1 interest	I.33, J.36, L.43, & F.23 (interesting texts, tasks, examples, & fun elements)
C.3 variety	F.22 (variety of topics from different fields)
C.4 grading	G.26 (efficiently graded tasks)
C.7 culture	F.18 (cultural sensitivities considered)
C.8 authenticity	G.27 (tasks: authentic/close to real language situations)
Grammar (D. Grammar, 1-4)	Grammar (L.42-45)
Vocabulary (D. Vocabulary, 1-2 & D. Vocabulary & structure 1-12)	Vocabulary (F.23-26)
Exercises and Activities (E.1-13)	Exercises (I.34-36) + Methodology (B.2-3)
Layout and physical makeup (F.1-9)	Physical and utilitarian attributes (D.8-13)

Table 1 shows the overlapping domains of the two checklists. The items in the Skierso checklist that were not in the ELT Textbook Evaluation Checklist included the domain of 'bibliographic data' and some of the items in 'subject matter', including C.2 'ordering', C.5 'level of abstractness', C.6 'register' and C.9 'cultural integration'. Excluding the items that did not match in the two checklists, the scores assigned for the textbook in reference to the remaining items (51 in the Skierso checklist and 35 in the ELT Textbook Evaluation Checklist, Table 1) were tested for their correlation.

The second instrument was the ELT Textbook Evaluation Checklist Evaluation Questionnaire [7]. This instrument is a five-point Likert style questionnaire in which a value of 1 signifies 'strongly disagree'; 2 'disagree'; 3 'not sure'; 4 'agree'; and 5 'strongly agree'. It is a concise questionnaire of 13 items focusing on the usefulness (that is, reliability, construct validity, impact and practicality) of the checklist. These items are followed by a final open-ended question which allows the evaluator to provide critical feedback on the checklist under examination. In order to collect data on the final research question; that is economy of the checklist, an item was added to this questionnaire which was, 'How much time did you spend to evaluate the textbook using this checklist?'

**Research Procedure:** First, the two evaluators were briefed on the objective of the study and the instruments that they were going to use. They were, then, given a copy of the ELT Textbook Evaluation Checklist Evaluation Questionnaire and were told that after using each checklist to evaluate the textbook, they were

supposed to evaluate the checklist itself. The evaluators' satisfaction of the checklist (the final research question) would be measured by analyzing the data elicited from the evaluators by the ELT Textbook Evaluation Checklist Evaluation Questionnaire. The evaluators were reminded to record the time that they spent evaluating the book using each instrument. The economy of the checklist would be measured by comparing the average of the time that the evaluators spent to evaluate the textbook using each checklist.

For data analysis purposes SPSS (Version 20) was used. Both descriptive and inferential statistics were used to analyze the data. The scores assigned by the two evaluators were correlated using the Pearson product-moment to test the concurrent validity as well as the inter-rater and internal reliability of the checklist.

## RESULTS

This section reports the findings of the study. The results include the checklist reliability (research question 1), concurrent validity (research question 2) and economy (research question 3), as well as the evaluators' levels of satisfaction with it (research question 4).

**Reliability:** The first research question dealt with the reliability of the checklist. More specifically, it focused on the inter-rater reliability of the ratings assigned by two evaluators using the checklist. To test the inter-rater reliability, two evaluators rated an English language teaching textbook, using the checklist. A Pearson product-moment correlation coefficient was computed to assess the correlation between their scores. According to

the results, there was a significant correlation between the two evaluators' scores ( $r = 0.88, n = 35, p = .000$ ). A correlation coefficient of below .50 is regarded as low, .50 to .75 as moderate and .75 to .90 as high [17]. Therefore, the strength of the correlation ( $r=0.88$ ) was 'high'. The results suggest that the checklist helped the two evaluators rate consistently. Additionally, another Pearson product-moment correlation coefficient was run to assess the correlation between the evaluators' scores using the Skierso checklist. As the results indicated, there was a significant correlation between the two teachers' scores ( $r = 0.76, n = 51, p = .000$ ). This correlation was high, too, but it was relatively lower than the inter-rater reliability achieved using the newly developed checklist. It can be concluded that the checklist could help evaluators rate textbooks more consistently than the Skierso checklist.

**Concurrent Validity:** The second research question concerned the concurrent validity of the checklist in reference to the Skierso checklist. When the scores produced by a newly developed instrument have a high correlation with those of another well-established instrument, it has high concurrent validity. To test the concurrent validity of the checklist, both the evaluators rated the same textbooks but this time using the Skierso checklist. A Pearson product-moment correlation coefficient was computed to assess the correlation between the scores produced by the two checklists. As mentioned before, only the items that overlapped in the two instruments (Table 1) were considered for analysis. The results indicated a moderate and significant correlation between the scores assigned by the first evaluator using the two checklists ( $r = 0.70, n = 51, 35, p = .000$ ). Regarding the second evaluator's scores, better results were obtained. A high and significant correlation was observed between the scores assigned by the second evaluator using the two checklists ( $r = 0.753, n = 51, 35, p = .000$ ).

**Economy:** The evaluators recorded the time that they spent evaluating the textbook using the checklist and the Skierso checklist. According to their records, on average, it took each evaluator 9.5 minutes to rate the textbook using the checklist. This was considerably shorter (almost three times) as compared to the time that they spent using the Skierso checklist (28.5 minutes).

**Evaluators' Satisfaction:** The evaluators' satisfaction of the checklist was determined using the ELT Textbook Evaluation Checklist Evaluation Questionnaire. The respondents' positive response to all the items (excluding three reverse items) would imply their satisfaction of the checklist. The reverse items would indicate how honest the evaluators had been in their responses. Positive responses to both a regular and a reverse item would imply that the respondent had failed to provide accurate and honest responses and all his other responses to the questionnaire would be regarded as unreliable and thus the questionnaire would be discarded. The two evaluators' responses to the questionnaire were examined and turned out to be accurate and thus suitable for analysis. A mean score of 5 would suggest that a given evaluator completely agreed that the checklist was 100% a useful instrument. In the case of the present study, mean scores of 4.46 and 4.15 were obtained from the first and the second evaluators, respectively. This suggested that overall they 'agreed' that the checklist was a useful instrument.

## DISCUSSION

The preceding section presented the results of the present study. Generally, the ELT Textbook Evaluation Checklist indicates high levels of reliability, validity and economy. Such promising results were not beyond the present researchers' expectation since this instrument has passed through several tests and has been refined based on their findings.

As the results of the first two research questions showed, the checklist indicated a high level of inter-rater reliability ( $r=.88$ ) as well as moderate ( $r = 0.70$ ) to high ( $r = 0.753$ ) levels of concurrent validity. These findings were consistent with those of our previous test results on the reliability and validity of the checklist. In a previous study, a group of English language teachers ( $n=82$ ) used the checklist to evaluate a textbook and then responded to the ELT Textbook Evaluation Checklist Evaluation Questionnaire to determine their perceptions towards the reliability and validity of the checklist [7]. According to the results, mean scores of 4.0 or more (upon a total of 5.0) were obtained for the reliability and validity of the checklist. This would suggest that the teachers who used the checklist in that study regarded it as a reliable and valid instrument. The findings of the present study provided empirical proof for the reliability and validity of the checklist.

As for the economy of the checklist, researchers [7] report that the ELT Textbook Evaluation Checklist (with 361 running words) is about 12 times shorter than the Skierso checklist (with 4553 running words). In this study, the findings of the third research question concerning the economy of the checklist verified the findings of the previous research conducted on the economy of the instrument. The evaluators spent considerably shorter time (9.5 minutes) to evaluate the same textbook as they spent using the Skierso checklist (28.5 minutes).

The evaluators' satisfaction levels of the checklist also corresponded with the teachers' satisfaction levels as reported by researchers[7]. In their study, the majority (78%) of the teachers indicated 'high' and more than one in ten (12.2%) reported 'very high' levels of satisfaction with the ELT Textbook Evaluation Checklist.

No responses were provided by any of the evaluators on the open-ended question at the end of the questionnaire. A possible reason could be that they focused more on evaluating the textbook than the checklist. Evaluators' fatigue could have also contributed to this problem. These two evaluators had to rate the same book using two different checklists and then use a questionnaire to evaluate one of the checklists. Therefore, the fact that they did not provide any responses for the open-ended question does not mean that they had, in fact, no opinions on how the checklist could be improved.

## CONCLUSION

Checklists are useful instruments that can help teachers and syllabus designers before, while and after using textbooks in a target learning-teaching context [18]. However, before they can be used confidently, these instruments should recursively undergo field-tests to ensure their validity. The present study was an attempt to evaluate the inter-rater reliability, concurrent validity and economy of a newly developed textbook evaluation instrument, called the ELT Textbook Evaluation Checklist. It also aimed at determining the extent to which the prospective users of the checklist would indicate they were satisfied with it.

Based on the findings of the present and previous related studies, it can be concluded that the ELT Textbook Evaluation Checklist may be used as a reliable, valid and practical checklist for evaluating English teaching textbooks. However, there are at least three points that should be closely taken into consideration before any generalizations could be made based on the findings of the present study:

- The users of the ELT Textbook Evaluation Checklist or any other similar instrument should not neglect the specific features of their testing situation. Variations in the test situation can put the relevance (or validity) of a checklist under question. Adaptations may be necessary if the checklist is going to be used for certain contexts. For example, if the objective is to evaluate the suitability of an English textbook for children, certain domains of the checklist (like grammar) may sound redundant and the textbook should not be rated in reference to the criteria that are not emphasized in it.
- In the present study, only two evaluators' scores were analyzed to test the reliability and validity of the checklist. Using more evaluators may result in different outcomes. Also, the evaluators had to be selected from a similar learning-teaching background and with almost the same level of experience. Further research that examines the checklist with the help of teachers from different teaching backgrounds and levels of experience may shed light on certain weaknesses of the checklist.
- The evaluators provided no responses for the open-ended question at the end of the questionnaire. Further research is needed in which a number of experts are asked for their critical comments on the adequacy, clarity and inclusiveness of the items in the checklist. This may shed light on the potential weaknesses that could have been overlooked by the two evaluators in this study.

The present paper can be useful for English language teachers material developers, syllabus designers and researchers in the area of ELT. The checklist can still be improved through progressive field-tests that may result in certain modifications to it. However, it can be used for evaluating ELT textbooks as an instrument that has been developed based on an extensive review of the literature and then been rigorously tested for its validity, reliability and practicality.

## REFERENCES

1. Skierso, A., 1991. Textbook Selection and Evaluation (2<sup>nd</sup> ed.). In M. Celce-Murcia. Teaching English as a Second or Foreign Language (pp. 432-453). Boston: Heinle and Heinle Publishers.
2. Sheldon, L.E., 1988. Evaluating ELT Textbooks and Materials. *ELT Journal*, 42(4): 237-246.

3. Mukundan, J. and T. Ahour, 2010. A Review of Textbook Evaluation Checklists across Four Decades (1970-2008). In B. Tomlinson and H. Masuhara (Eds.), *Research for Materials Development in Language Learning: Evidence for Best Practice* (pp: 336-352). London: Continuum.
4. Mukundan, J., R. Hajimohammadi and V. Nimehchisalem, 2011. Developing an English Language Textbook Evaluation Checklist. *Contemporary Issues in Education Research*, 4(6): 21-27.
5. Mukundan, J., V. Nimehchisalem and R. Hajimohammadi, 2011. Developing an English Language Textbook Evaluation Checklist: A Focus Group Study. *International Journal of Humanities and Social Science*, 1(12): 100-105.
6. Mukundan, J. and V. Nimehchisalem, XXXX. In Press. Evaluative criteria of an English language textbook evaluation checklist. Accepted to be published in the *Journal of Language Teaching and Research*.
7. Nimehchisalem, V. and J. Mukundan, XXXX. In Press. Usefulness of the English Language Teaching Textbook Evaluation Checklist. *Pertanika Journal of Social Sciences and Humanities (JSSH)*.
8. Weir, C. J., 2005. *Language Testing and Validation: An Evidence-based Approach*. Hampshire, UK: Palgrave MacMillan.
9. Lado, R., 1961. *Language Testing*. London: Longman.
10. Messick, S., 1989. Validity. In R. L. Linn (Ed.), *Educational Measurement*, New York: Macmillan.
11. Tomlinson, B., (Ed.), 1998. *Materials Development in Language Teaching*. Cambridge: Cambridge University Press.
12. Stevick, E.W., 1971. *Adapting and Writing Language Lessons*. Washington, DC: Government Printing Office.
13. Brown, J.D., 1995. *The Elements of Language Curriculum*. Boston: Heinle and Heinle Publishers.
14. Ansary, H. and E. Babaii, 2002. Universal Characteristics of EFL/ESL Textbooks: A Step towards Systematic Textbook Evaluation. *The Internet TESL Journal*, VIII (2) Retrieved on March 1, 2012 from <http://iteslj.org/Articles/Ansary-Textbooks/>
15. Soars, L. and J. Soars, 1996. *New Headway* (2<sup>nd</sup> ed.). Oxford: Oxford University Press.
16. Mukundan, J., 2009. *ESL Textbook Evaluation: A Composite Framework*. Köln, Germany: Lambert Academic Publishing.
17. Farhadi, H., A. Jafarpur and P. Birjandi, 2001. *Testing Language Skills: From Theory to Practice*. Tehran, Iran: SAMT.
18. Ellis, R., 1997. *SLA Research and Language Teaching*. Oxford: Oxford University Press.