

Evaluative Criteria of an English Language Textbook Evaluation Checklist

Jayakaran Mukundan (Corresponding author)

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

Email: jayakaranmukundan@yahoo.com

Vahid Nimehchisalem

Department of Educational Studies, Faculty of Educational Studies, Universiti Putra Malaysia, Malaysia

Email: nimechie22@yahoo.com

Abstract—Checklists are instruments that help teachers or researchers in the area of English Language Teaching (ELT) to evaluate teaching-learning materials like textbooks. Several checklists are available in the literature, most of which lack validity. The paper discusses the results of a survey that investigated a group of English as a Second Language (ESL) experts' (n=207) views on a checklist developed by the present researchers. The results showed an equal level of importance for all the items of the checklist. Additionally, based on the findings of factor analysis, two items were removed from the checklist. The study offers useful implications for ELT practitioners and researchers. Further research is necessary to field test the checklist for its validity and reliability.

Index Terms—English language teaching material evaluation, textbook evaluation checklists

I. INTRODUCTION

The textbook is one of the crucial factors in determining the learners' success in language courses. Teachers or curriculum developers, therefore, should select this teaching material carefully. Checklists are often used by experts in evaluating and selecting textbooks. Evaluation is made easier, more objective and valid when it is based on a reliable instrument. Most checklists available in the literature lack the expected validity or reliability (Mukundan & Ahour, 2010). This necessitates the need for developing a checklist that is of high validity in terms of the construct domain of its evaluative criteria, that accounts for the consistency of the scores resulting from its items, and that is economical.

This study presents part of a project, the objective of which was to develop the English Language Teaching Textbook Evaluation Checklist (ELT-TEC). The project commenced by a review of the available instruments (Mukundan & Ahour, 2010). In the light of the evaluative criteria in the available well-established checklists, the researchers developed a tentative checklist (Mukundan, Hajimohammadi, & Nimehchisalem, 2011). This was followed by a qualitative study in which a focus group, including six ELT experts, helped the researchers enhance the clarity and inclusiveness of the checklist (Mukundan, Nimehchisalem, & Hajimohammadi, 2011). Parallel with the focus group, a survey of a group of English as a Second Language (ESL) experts' views on the tentative checklist was conducted. The present paper reports the findings of this survey.

In developing checklists, two important matters stand out. One of them is determining the evaluative criteria that constitute the main skeleton of any checklist and according to which textbooks are evaluated. The other crucial step is to decide on the level of importance or 'weight' of each criterion. This paper was mainly concerned with the second issue. With regard to their weightage, checklists may be of equal-weight or optimal-weight schemes. In an equal-weight scheme equal weights are assigned to each criterion whereas in an optimal-weight scheme different weights are assigned to each criterion. Most checklists available in the literature follow an equal weight scheme.

II. OBJECTIVES AND RESEARCH QUESTIONS

The objective of the study was to determine the degree of importance of each section alongside its related sub-categories. The study also sought to test the practical significance of each item. The following research questions were addressed:

1. What is the degree of importance of each section and sub-category of the checklist?
2. Which items should be included in the final checklist?

III. METHOD

Quantitative method was used for collecting and analyzing the data. This section discusses the sample, instrument, and data analysis method used in the study.

A. Sample

The sample included 207 English language teachers or lecturers (72.5% female, aged between 20 and 67) in Malaysia.

B. Instrument

The instrument that was administered was a revised version of the tentative checklist for textbook evaluation (Mukundan, Hajimohammadi, & Nimehchisalem, 2011). As presented in Appendix I, the checklist was converted into a 5-scale Likert style questionnaire which consisted of two parts. The first part elicited demographic information, like the respondents' gender, age, level of education, teaching context, as well as teaching experience. It also inquired whether the respondents had experienced any workshops related to textbook evaluation or selection and whether they had already been involved in textbooks evaluation or selection.

The second part presented the checklist to the respondents and asked them to read and rate the importance of each item from 0 (for unimportant) to 4 (very important). The respondents were told that they could add or delete sections, sub-categories, or items based on their own judgement. The second part of the questionnaire also provided two additional columns in front of each section, sub-category or item. In the first column, they could reword any part of this section or comment on it whenever they regarded it fit. The 'reword' and 'comment' columns would enable the researchers to collect some more qualitative data to support the findings or the focus group study (Mukundan et al., 2011).

C. Data Analysis

The collected data were analyzed using SPSS version 16. Descriptive statistics and Exploratory Factor Analysis were used for analyzing the data. Factor analysis can indicate how much variance is explained by each factor, or in the case of the present study, each sub-category (e.g., methodology, suitability to learners, physical and utilitarian attributes, etc.). Moreover, it can help instrument developers in grouping several items under a limited number of categories (Hair, Black, Babin, Anderson, & Tatham, 2006). Such an application can prove very useful in developing instruments since it can help developers come up with a more economical instrument by collapsing certain components.

IV. RESULTS AND DISCUSSION

Table 1 shows the descriptive statistics related to the demographics of the respondents (n=207), who were 72.5% females, aged between 20 and 67.

TABLE I.
DESCRIPTIVE STATISTICS RESULTS OF THE DEMOGRAPHIC QUESTIONNAIRE

Demographic feature	Category	Frequency	Percentage (%)
Gender	Male	57	27.5
	Female	150	72.5
Age	20-35 years of age	112	54.1
	36-50 years of age	83	40.1
	51-67 years of age	12	5.8
Level of education	Diploma	20	9.7
	BA	107	51.7
	MA	75	36.2
	PhD	5	2.4
Teaching context	School	83	40.1
	Language institute	25	12.1
	University	99	47.8
Teaching experience	1-5 years (low)	82	39.6
	6-15 (moderate)	71	34.3
	16-35 (high)	54	26.1

As for their level of education, more than half of the respondents (57.7%) held a bachelor degree (57.7%). The remaining part had a diploma (9.7%), master degree (36.2%), or PhD (2.4%). Most of the respondents taught at university (47.8%). Another large proportion (40.1%) of the respondents comprised school teachers. Language institute instructors constituted the smallest group (12.1%). The teaching experience of the respondents ranged between 1 and 35 years, with a majority (39.6%) having a low and a minority (26.1%) having a high teaching experience.

The questionnaire was administered to the respondents. The survey resulted in quantitative data that were analyzed using descriptive statistics and factor analysis methods, the results of which are discussed in this section. In order to answer the first research question; that is, the importance of each criterion, the mean and percentage of each criterion as rated by the respondents were calculated (Table 2).

TABLE II.
DESCRIPTIVE STATISTICS RESULTS FOR THE IMPORTANCE OF CRITERIA

No	Criteria	n	Sum	Mean	Std.*	%
1.	The book in relation to syllabus	207	659	3.18	.73	7.8
2.	Methodology	207	639	3.09	.65	7.6
3.	Suitability to learners	207	640	3.09	.68	7.6
4.	Physical and utilitarian attributes	207	626	3.02	.64	7.4
5.	Supplementary materials	207	651	3.15	.79	7.7
6.	Listening	207	642	3.10	.75	7.6
7.	Speaking	207	661.5	3.19	.72	7.8
8.	Reading	207	655	3.16	.73	7.8
9.	Writing	207	651.5	3.15	.79	7.7
10.	Vocabulary	207	662	3.20	.66	7.9
11.	Grammar	207	645	3.12	.66	7.7
12.	Pronunciation	207	636	3.07	.72	7.6
13.	Exercises	207	656	3.17	.70	7.8

*Std.: Standard deviation

The results are expressed as mean \pm SD ($n = 207$). As it can be observed from the table, 'vocabulary' was rated as the most important ($3.2 \pm .66$) and 'physical and utilitarian attributes' as the least important ($3.02 \pm .64$) criteria. This would mean a range of 7.4% and 7.9%; that is, an inconsiderable difference of 0.5% between the variables with the highest and lowest degree of importance. This finding suggests that the respondents regarded all the criteria as either 'important' (3.0) or 'very important' (4.0). Therefore, it is not necessary to assign weights for the criteria in the checklist. Such findings are in line with the literature in which most checklists are of an equal-scheme weight.

Factor analysis was used to analyze the data and provide an answer for the second question, the significance of each item. There are two points that should be considered before using factor analysis, the sample size and strength of inter-correlations among items. The appropriate sample size for factor analysis is a size of 10:1 (or a minimum of 5:1) ratio of respondents to items (Nunnally, 1978). The instrument had 38 items, so a size ranging between 190 and 380 would be appropriate for this study. The sample size of the present study, 207, falls within this range and is therefore suitable for factor analysis. Another criterion to check the adequacy of the sample size is Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy. If its value exceeds the threshold of .6, the sample size is adequate. Table 3 shows the SPSS output for KMO test.

TABLE III.
KMO AND BARTLETT'S TEST

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.931
Bartlett's Test of Sphericity	Approx. Chi-Square	6.116E3
	df	703
	Sig.	.000

According to the table, the calculated value of KMO measure is .931 that is more than .6, which suggests the adequacy of the sample size. As for the strength of inter-correlations among items, based on the result of Bartlett's test of sphericity (Table 3), the significant value is smaller than alpha at .05 level of significance ($p = .000 < \alpha = .05$), which suggests that the data set is suitable for factor analysis.

The Varimax rotation technique was used to determine the factor loading of each item. Appendix II shows the results of this analysis. The rotated component matrix (Appendix II) can be used for two purposes. First, it helps researchers group different items under certain categories. Second, it can indicate which items are practically significant. With regard to grouping the items, since the results of factor analysis were not consistent with the literature, they were not followed in this study. The results were used for determining the level of significance of each item.

In Appendix II, the values in front of each component, or item, are called factor loadings that show the correlation between the original variables and the factors (Coakes & Steed, 2007). Once the factor loadings are squared, they can indicate the percentage of variance in an original variable explained by a factor. Appendix III shows the squared factor loading of each item. The results were interpreted based on Hair et al's (2006) rule of thumb:

Unacceptable factor: $>.30$

Minimally acceptable factor: $.30-.40$

Acceptable factor: $.40-.50$

Significant factor $.50 <$

As Hair *et al.* (2006) also point out, researchers' final decision on the number of the factors to be included in the final instrument should rely on the literature. Factor analysis results should only be regarded as recommendations. Table 4 presents the items categorized following Hair et al's (2006) rule of thumb.

TABLE IV.
ITEMS CATEGORIZED BASED ON HAIR ET AL'S (2006) RULE OF THUMB

Unacceptable items (>.30)	Minimally acceptable items (.30-.40)	Acceptable items (.40-.50)	Significant items (.50<)
6, 11, 12, 15, 27	3, 4, 5, 7, 14, 16, 17, 18, 20, 28, 29, 30, 31, 32, 38	1, 10, 19, 21, 22, 23, 24, 25, 26, 33, 34, 35, 36	2, 8, 9, 13, 37

According to the table, 5 items were regarded as unacceptable because their squared factor loadings were less than .30. These items included 'the compatibility to the learner needs' (item 6), 'efficient audio materials' (item 11), 'interesting tasks' (item 12), 'cultural sensitivities' (item 15), and 'vocabulary load' (item 27). According to the literature, these items are all important and should not be removed from the final checklist; however, a closer look at the items in the checklist shows their redundancy. As for item 6, there are already two other items (4 and 5) that consider the compatibility of the checklist to the learners' age and needs. Including another item would therefore sound unnecessary. The same argument also seems true for item 12 since the same feature has been repeated in items 24 (Texts are interesting), 26 (Tasks are interesting), and 32 (Examples are interesting). Therefore, it seems logical to remove this item based on the results of factor analysis. It may be argued that removing these items may lower the reliability of the instrument and that keeping them in the final checklist will not reduce its validity. It should, however, be noted that removing the unnecessary items will result in the higher economy of the instrument and enhance its usefulness.

Items 11, 15 and 27 were not removed from the final checklist, however. The reason was that regarding the literature, the researchers considered 'efficient audio materials' (item 11), 'cultural sensitivities' (item 15), and 'vocabulary load' (item 27) as important items. Removing these items would affect the construct validity of the checklist.

V. CONCLUSION

The paper presented the findings of the quantitative phase of a project that aims at developing a checklist to evaluate English language teaching textbooks. The main objective of the present study was to provide proof for the construct validity of the checklist. The results showed that the respondents viewed all the items as equally important. Two of the items were found to be redundant and were removed from the final checklist to improve its economy.

This study indicates how quantitative method can be employed to provide support for the validity of instruments in their development process. As it was also observed, factor analysis can help developers in making their instrument more economical. Despite their usefulness, factor analysis results must be handled cautiously and should be interpreted in the light of the related literature.

ESL researchers and teachers and particularly ELT material developers and evaluators will find the results of this research useful. The study also provides a practical guide for curriculum developers and indicates what constitutes a good textbook based on the views of ELT practitioners.

The findings of the present study helped the researchers further refine their checklist, which at this stage can be applied for evaluating textbooks more confidently. Future study will focus on the empirical test of the reliability and validity of the checklist.

APPENDIX I. ENGLISH LANGUAGE TEACHING TEXTBOOK CHECKLIST

Dear respondent

This project aims at finding out what evaluative criteria are important for English language teachers or lecturers. Please answer the following questionnaire regarding your personal and professional background.

- Gender: ☐ Male ☐ Female
- Age: years
- Level of education: ☐ Diploma ☐ BA ☐ MA ☐ PhD
- Major: ☐ TESL ☐ Others
- Teaching context: ☐ University ☐ School ☐ Language institute
- Teaching experience:years
- Have you ever participated in any textbook evaluation workshops, seminars, courses, etc.?
☐ Yes ☐ No ☐ Not applicable
- If your answer to question 7 is 'Yes', please list the courses you attended.
a)
b)
c)
- Have you ever evaluated a textbook? ☐ Yes ☐ No

In the following section, you will find a list of the criteria that will be used to develop a checklist for evaluating English language teaching textbooks. You are requested to mark (0-4) to indicate the level of importance of each criterion according to this key:

0: Unimportant 1: Less important 2: Fairly important 3: Important 4: Very important

If you think a criterion is missing, you may add it to the end of the list and indicate its level of importance. In addition, if there is a term that, according to your experience, would be hard for evaluators to understand, you may add the term that you recommend in the column, *Reword*. If you have any further comments about each criterion, you may mention it in the *Comment* column.

Thank you for your cooperation.

ENGLISH LANGUAGE TEACHING TEXTBOOK CHECKLIST

Evaluative criteria	Level of importance	Reword	Comment
I. General attributes			
A. The book in relation to syllabus and curriculum			
1. It matches to the specifications of the syllabus.	① ② ③ ④		
B. Methodology			
2. The activities can be exploited fully and can embrace the various methodologies in ELT.	① ② ③ ④		
3. Activities can work well with methodologies in ELT.	① ② ③ ④		
C. Suitability to learners			
4. It is compatible to the age of the learners.	① ② ③ ④		
5. It is compatible to the needs of the learners.	① ② ③ ④		
6. It is compatible to the interests of the learners.	① ② ③ ④		
D. Physical and utilitarian attributes			
7. Its layout is attractive.	① ② ③ ④		
8. It indicates efficient use of text and visuals.	① ② ③ ④		
9. It is durable.	① ② ③ ④		
10. It is cost-effective.	① ② ③ ④		
E. Efficient outlay of supplementary materials			
11. The book is supported efficiently by essentials like audio-materials.	① ② ③ ④		
II. Learning-teaching content			
A. General			
12. Most of the tasks in the book are interesting.	① ② ③ ④		
13. Tasks move from simple to complex.	① ② ③ ④		
14. Task objectives are achievable.	① ② ③ ④		
15. Cultural sensitivities have been considered.	① ② ③ ④		
16. The language in the textbook is natural and real.	① ② ③ ④		
17. The situations created in the dialogues sound natural and real.	① ② ③ ④		
B. Listening			
18. The book has appropriate listening tasks with well-defined goals.	① ② ③ ④		
19. Tasks are efficiently graded according to complexity.	① ② ③ ④		
20. Tasks are authentic or close to real language situations.	① ② ③ ④		
C. Speaking			
21. Activities are developed to initiate meaningful communication.	① ② ③ ④		
22. Activities are balanced between individual response, pair work and group work.	① ② ③ ④		
D. Reading			
23. Texts are graded.	① ② ③ ④		
24. Texts are interesting.	① ② ③ ④		
E. Writing			
25. Tasks have achievable goals and take into consideration learner capabilities.	① ② ③ ④		
26. Tasks are interesting.	① ② ③ ④		
F. Vocabulary			
27. The load (number of new words in each lesson) is appropriate to the level.	① ② ③ ④		
28. There is a good distribution (simple to complex) of vocabulary load across chapters and the whole book.	① ② ③ ④		
29. Words are efficiently repeated and recycled across the book.	① ② ③ ④		
G. Grammar			
30. The spread of grammar is achievable.	① ② ③ ④		
31. The grammar is contextualized.	① ② ③ ④		
32. Examples are interesting.	① ② ③ ④		
33. Grammar is introduced explicitly and reworked incidentally throughout the book.	① ② ③ ④		
H. Pronunciation			
34. It is contextualized.	① ② ③ ④		
35. It is learner-friendly with no complex charts.	① ② ③ ④		
I. Exercises			
36. They are learner friendly.	① ② ③ ④		
37. They are adequate.	① ② ③ ④		
38. They help students who are under/over-achievers.	① ② ③ ④		
39.			
40.			

APPENDIX II. ROTATED COMPONENT MATRIX^A

	Component					
	1	2	3	4	5	6
II.D. Texts are interesting	.693					
II.C. Activities are developed to initiate meaningful communication	.675					
II.D. Texts are graded	.663					
II.E. Tasks have achievable goals and take into consideration learner capabilities	.661					
II.E. Tasks are interesting	.656					
II.C. Activities are balanced between individual response, pair work and group work	.635					
II.F. There is a good distribution (simple to complex) of vocabulary load across chapters and the whole book	.552					
II.F. The load (number of new words in each lesson) is appropriate to the level	.543					
II.I. They are adequate		.715				
II.H. It is learner-friendly with no complex charts		.679				
II.H. It is contextualized		.656				
II.I. They are learner friendly		.654				
II.G. Grammar is introduced explicitly and reworked incidentally throughout the book		.654				
II.G. The spread of grammar is achievable		.579				
II.G. The grammar is contextualized		.570				
II.I. They help students who are under/over-achievers		.565				
II.F. Words are efficiently repeated and recycled across the book		.551				
II.G. Examples are interesting		.549				
I.B. The activities can be exploited fully and can embrace the various methodologies in ELT			.734			
I.A. It matches to syllabus specifications			.688			
I.B. Activities can work well with methodologies in ELT			.626			
I.C. It is compatible to the needs of the learners			.596			
I.C. It is compatible to the age of the learners			.569			
I.C. It is compatible to the interests of the learners			.464			
II.B. Tasks are efficiently graded according to complexity				.638		
II.A. The situations created in the dialogues sound natural and real				.629		
II.B. The book has appropriate listening tasks with well-defined goals				.621		
II.A. The language in the textbook is natural and real				.611		
II.B. Tasks are authentic or close to real language situations				.560		
I.E. The book is supported efficiently by essentials like audio-materials				.537		
II.A. Cultural sensitivities have been considered				.412		
I.D. It is durable					.781	
I.D. It indicates efficient use of text and visuals					.724	
I.D. It is cost-effective					.660	
I.D. Its layout is attractive					.623	
II.A. Tasks move from simple to complex						.726
II.A. Task objectives are achievable						.625
II.A. Most of the tasks in the book are interesting						.534

Extraction Method: Principal Component Analysis.
 Rotation Method: Varimax with Kaiser Normalization.
 a. Rotation converged in 10 iterations.

APPENDIX III. SQUARED FACTOR LOADINGS

Item	Factor loading	Squared factor loading
Section I		
1. It matches to the specifications of the syllabus.	.688	.473
2. The activities can be exploited fully and can embrace the various methodologies in ELT.	.734	.539
3. Activities can work well with methodologies in ELT.	.626	.392
4. It is compatible to the age of the learners.	.569	.324
5. It is compatible to the needs of the learners.	.596	.355
6. It is compatible to the interests of the learners.	.464	.215
7. Its layout is attractive.	.623	.388
8. It indicates efficient use of text and visuals.	.724	.524
9. It is durable.	.781	.700
10. It is cost-effective.	.660	.436
11. The book is supported efficiently by essentials like audio-materials.	.537	.288
Section II		
12. Most of the tasks in the book are interesting.	.534	.285
13. Tasks move from simple to complex.	.726	.527
14. Task objectives are achievable.	.625	.391
15. Cultural sensitivities have been considered.	.412	.170
16. The language in the textbook is natural and real.	.611	.373
17. The situations created in the dialogues sound natural and real.	.629	.396
18. The book has appropriate listening tasks with well-defined goals.	.621	.386
19. Tasks are efficiently graded according to complexity.	.638	.407
20. Tasks are authentic or close to real language situations.	.560	.314
21. Activities are developed to initiate meaningful communication.	.675	.456
22. Activities are balanced between individual response, pair work and group work.	.635	.403
23. Texts are graded.	.663	.440
24. Texts are interesting.	.693	.480
25. Tasks have achievable goals and take into consideration learner capabilities.	.661	.437
26. Tasks are interesting.	.656	.430
27. The load (number of new words in each lesson) is appropriate to the level.	.543	.294
28. There is a good distribution (simple to complex) of vocabulary load across chapters and the whole book.	.552	.305
29. Words are efficiently repeated and recycled across the book.	.551	.304
30. The spread of grammar is achievable.	.579	.335
31. The grammar is contextualized.	.570	.325
32. Examples are interesting.	.549	.301
33. Grammar is introduced explicitly and reworked incidentally throughout the book.	.654	.428
34. It is contextualized.	.656	.430
35. It is learner-friendly with no complex charts.	.679	.461
36. They are learner friendly.	.654	.428
37. They are adequate.	.715	.511
38. They help students who are under/over-achievers.	.565	.319

REFERENCES

- [1] Coakes, S. J. & L. Steed (2007). SPSS version 14.0 for Windows: Analysis without anguish. Melbourne: Wiley.
- [2] Hair, J. F., W. C. Black, B. J. Babin, R. E. Anderson, & R. L. Tatham. (2006). Multivariate data analysis (6th edn.). NJ: Pearson.
- [3] Mukundan, J. & T. Ahour. (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*. London: Continuum, 336-352.
- [4] Mukundan, J., R. Hajimohammadi, & V. Nimehchisalem. (2011). Developing an English language textbook evaluation checklist. *Contemporary Issues in Education Research*. 4.6, 21-27.
- [5] Mukundan, J., V. Nimehchisalem, & 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] Nunnally, J. C. (1978). *Psychometric theory* (2nd edn.). New York: McGraw-Hill.

Jayakaran Mukundan is a Professor at the Faculty of Educational Studies, UPM, Malaysia and a Visiting Fellow at Leeds Metropolitan University, UK. He is also a Director on the Extensive Reading Foundation Board. His areas of interest include ESL writing and ELT textbook evaluation.

Vahid Nimehchisalem has been an English teacher, material developer, test developer, teacher trainer and lecturer. His main areas of research interest include assessing writing and ELT material evaluation. He completed his PhD in Teaching English as a Second Language (TESL) and, as a postdoctoral fellow, is now involved in research projects on ELT at the Faculty of Educational Studies, UPM, Malaysia.