Developing a Checklist for Evaluating the Presentation of New Vocabulary in ELT Textbooks

Nur Izyan Syamimi Mat Hussin, Vahid Nimehchisalem[*],

Seyed Ali Rezvani Kalajahi

Department of English Language, Faculty of Modern Languages and Communication,
Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.

Abstract

Textbook evaluation can be a guide and reference to the educational institutions, curriculum designers, writers and educators in selecting the suitable material regarding their learners' needs. A checklist is one of the common instruments used in textbook evaluation studies. Many checklists have been developed to evaluate whole textbooks. However, few checklists are available that focus on a specific feature of an English language learning-teaching textbook. This study presents the development procedure of a checklist that focuses on the presentation of vocabulary in ELT textbooks. The instrument was developed in the light of a number of previous evaluation checklists. It was further refined based on interviews with a number of English language teachers, who commented on the relevance and clarity of its items. The reliability test revealed very high overall internal reliability of the instrument. The paper presents the checklist that can be used as a guide and reference to other researchers, curriculum designers, writers, teachers, as well as material evaluators.

© 2015 Penerbit Universiti Malaysia Pahang

Keywords: Checklist development; Checklist for vocabulary items; Textbook evaluation

INTRODUCTION

Textbooks are regarded as the main sources of learning and teaching English language in schools. In some countries, English textbooks are prescribed by Ministry of Education as a centralized decision-making agent (Byrd, 2001). In such systems, a unified series of textbooks is created for use throughout the country rather than selecting textbooks from a generic collection created by commercial publishing companies (Yap, 2011).

Based on Tomlinson (1998), learning-teaching material evaluation is "the systematic appraisal of the value of materials in relation to their objectives and to the objectives of the learners using them" (p.11). Evaluation can be in three forms; pre-use, in-use and post-use evaluation (Cunningsworth, 1995). Pre-use or predictive evaluation is designed to examine the future or potential performance of a textbook. In-use evaluation examines the material currently in use; and post-use or retrospective evaluation is concerned with the evaluation of textbooks after they have been used in a specific institution or situation (Ellis, 1997).

Checklists are commonly used by researchers to evaluate textbooks. A checklist can be used to evaluate the suitability and practicality of a textbook. Mukundan, Nimehchisalem and Hajimohammadi, (2011), AbdelWahab (2013), Miekley, (2005) as well as Daoud and Celce-Murcia (1979) are some of the researchers who used checklists to evaluate textbooks. Mukundan and Ahour (2010) in their review of textbook evaluation checklists within four decades (1970-2000) reported that most of the checklists developed by most researchers are qualitative. Most of the checklists have been developed to evaluate all of the features of a textbook, such as vocabulary, grammar, exercises, and pronunciation. Few focus primarily on a specific feature.

Textbooks play an important role in English language development of Malaysian school students. The textbook must be able to guide learners in enhancing their vocabulary skill and teachers in teaching their students what is necessary for students to master the new vocabulary effectively. It is necessary to study the quality of presentation of new vocabulary in textbooks, particularly when the textbook is the only resource for both learners and teachers in an English teaching-learning context. Therefore, there is a need for developing checklists for evaluating the presentation of new vocabulary items in textbooks. The current study was an attempt to develop a checklist for evaluating the presentation of new vocabulary items in Malaysian school English language textbooks.

1.1   Objectives and Research Questions

The objectives of the study were to develop a reliable, comprehensive, and clear checklist for evaluating the presentation of new vocabulary in ELT textbooks. The following research questions were posed to address these objectives:

1. What evaluative criteria should be included in the checklist?
2. To what extent do the items of the checklist indicate internal reliability?

1.2   Limitations and Scope

The study was limited only to the new vocabulary presentation in the Malaysian Secondary School English language textbooks. The checklist was refined through interviews with a limited number of Secondary School English teachers in Malaysia. Therefore, the checklist may not be appropriate for all types of ELT textbooks; for example, Primary school textbooks.

LITERATURE REVIEW

Many studies on textbook evaluation have been conducted to improve and modify textbooks. Many researchers have addressed the presentation of vocabulary in most textbook evaluation studies. For instance, Skierso (1991) constructed five sections in her checklist: bibliographical data, aims and goals, subject matter, vocabulary and structures, and layout and physical makeup. In another checklist, Miekley (2005) proposed four sections: content, vocabulary and grammar, exercises and activities, and attractiveness of the text and physical make-up. The vocabulary and grammar section of this checklist had five items, including (i) the presentation of the grammar rules in the textbook, (ii) various ways of the presentation the new vocabulary, (iii) appropriateness of the new vocabulary, (iv) repetition of the new vocabulary in the textbook, and (v) techniques in teaching-learning new vocabulary. Moreover, Mukundan, Nimehchisalem and Hajimohammadi (2011) in their attempt to develop a valid, reliable and practical checklist, allocated a section for vocabulary with four items. The first item was about the appropriateness of the load of the new words in each lesson of the textbook. The second dealt with the distribution of simple to complex of vocabulary load across the chapter and the whole book. The third item focused on the repetition and recycling of words across the book. The last item dealt with the contextualization of the words in the textbook.

Daoud and Celce-Murcia (1979) constructed nine items for the vocabulary and structure section of their checklist, including the following features:

1. reasonability of the load of new words introduced in every lesson;
2. repetition of the new vocabulary in subsequent lessons;
3. systematic gradation from simple to complex;
4. reasonability of the length of the sentences for the students' level;
5. appropriateness of the grammatical point and their sequence;
6. complexity of the structure;
7. use of current everyday language and sentence structures that follow normal order;
8. logical sequence of the sentence and paragraph; and
9. linguistic items introduced in meaningful situations.

Only four of these items were related to the presentation of vocabulary in the textbooks. AbdelWahab (2013) also developed an English language textbook evaluation checklist, by refining the previously developed checklists. He constructed 11 items under the vocabulary category which dealt with the appropriateness of the load of new words in each lesson, distribution of the vocabulary load across the whole book, vocabulary exercises, contextualization of the vocabulary, topical nature of the vocabulary exercises, new lexical items in each unit, specific methods used in teaching the new vocabulary, words used in sentences and examples, a list of vocabulary items at the end of the textbook, culture-specific items of the content and phonetic transcription of the words.

Most of the previous checklists do not thoroughly meet the requirements of a good and applicable instrument for evaluative purposes because they are either too short or too long and some criteria are vague (AbdelWahab, 2013). Another short-coming of the previous checklists is that they lack unidimensionality. Frazer and Lawley (2000) argue that instrument developers should ensure that each item addresses only one subject; otherwise, they will confuse the respondents. Arksey and Knight (1999, as cited in Gray, 2009) also discussed the same principle in their list of what to avoid when constructing individual questions. They stated that double-barreled questions should be avoided as they are impossible to answer. The answer provided by the respondents would be unclear as whether it relates to either items or only one of them. To offer a few examples for checklist items that violate the principle of unidimensionality, in Doud and Celce-Murcia (1979) checklist, there are two constructs in the first item, including 'vocabulary load' and 'repetition of vocabulary'. Likewise, the second items in the vocabulary sections of both Mukundan et al. (2011) and AbdelWahab (2013) checklists, *There is a good distribution (simple to complex) of vocabulary load across chapter and the whole book*, more than one construct is being addressed. In the development of the present checklist, the important notion of unidimensionality was considered and as it will be reported later, the aforementioned item was split into two separate items: *There is balance of simple to complex words* (Item 2) and *Good distribution of vocabulary load across the whole book* (Item 3).

Based on the review of textbook evaluation checklists within four decades (1970-2000), Mukundan and Ahour (2010) reported that most of the checklists are qualitative (e.g., Breen & Candlin, 1987; Bruder, 1978; Cunningsworth, 1984; Cunningsworth, 1995; Cunningsworth & Kusel, 1991; Dougill, 1987; Driss, 2006; Garinger, 2001; Garinger, 2002; Griffiths, 1995; Harmer, 1991; Haycraft , 1978; Hemsley, 1997; Hutchinson & Waters, 1987; Krug, 2002; Matthews, 1985; McDonough & Shaw, 2003; McGrath, 2002; Rahimy, 2007; Richards, 2001; Robinett , 1978; Rubdy, 2003; Sheldon, 1988; Zabawa, 2001) than quantitative (e.g., Canado & Esteban, 2005; Daoud & Celce-Murcia, 1979; Grant, 1987; Harmer, 1998; Litz, 2005; Miekley, 2005; Peacock, 1997; Sheldon, 1988; Skierso, 1991; Tucker, 1975; Ur, 1996; Williams, 1983;) or have head words/outline format, i.e., those without rating scales or questions (Ansari & Babaii, 2002; Brown, 1995; Littlejohn, 1998; Roberts, 1996). As it will be discussed, the review of the literature guided the developing process of the current checklist. It indicates that it is necessary to develop a new checklist that is suitable to examine the appropriateness of the vocabulary items in the Malaysian Secondary School English language textbooks.

METHODOLOGY

This developmental study followed a qualitative method to develop the checklist and refine it. Both qualitative and quantitative methods were used to finalize the checklist and to test its reliability, respectively.

3.1   Participants

The participants of this study were five Malaysian English Secondary School teachers. They were all female, aged between 30 and 60 years. The teachers had been involved in teaching English for more than ten years, and they had been using the English Secondary textbooks since it was introduced by the Ministry of Education in Malaysia in 2004. Therefore, they were well-aware of the characteristics of a good textbook and able to help the researchers with the construction of the checklist. They also had the experience of selecting and evaluating the English textbook and teaching materials.

3.2   Research Procedure

The process of developing the current checklist went through three phases:

Developing the prototype checklist
based on the literature

(Phase 1)

Refining the checklist by
interviewing teachers
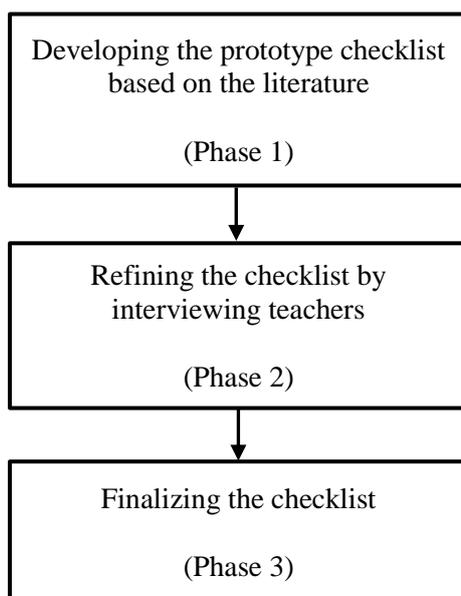
(Phase 2)

Finalizing the checklist

(Phase 3)

Figure 1. The development stages of the checklist

Figure 1 illustrates the three phases of developing checklist: developing the prototype checklist based on the literature, refining the checklist by interviewing the teachers and finalizing the checklist. During phase one, the literature and previous studies guided the process of development of the new vocabulary items in the checklist. The development of the checklist followed a set of ground rules which will be presented in the following section. In the second phase, the prototype checklist was presented to the teachers who gave their opinions and suggestions concerning the items during the semi-structured interview. The results from the interview were reported. Some items were deleted, split and/or modified. In the final stage, 11 items were finalized for evaluating the presentation of new vocabulary. Also, the overall internal reliability of the checklist and each item was tested using SPSS (Version 16). Due to its developmental nature, further details about the procedure of the study will be presented together with the results in the following section.

FINDINGS

4.1   Phase One

Based on Ary, Jacobs, and Razavieh (2002), the following ground rules were set to develop the checklist:

1. Avoid "double-barreled" items, which ask two questions in one item.
2. Avoid leading questions, which imply a desired response.
3. Avoid biased items that may predetermine a respondent's answer.
4. Avoid using ambiguous words and phrases.
5. Avoid items that might mislead because of unstated assumptions.
6. Make sure all the items are relevant to the issue.
7. Each item should be short, simple, and direct.
8. Number the items consecutively.
9. Use numbers to identify the various response categories.
10. Be consistent in assigning numbers to the various answer categories.
11. Response categories should be arranged vertically rather than horizontally.

The current checklist was developed based on the review of related literature and previous studies of developing English language textbook evaluation checklists. Appendix A shows the prototype checklist and the source of each of its items. The items follow their original sequences. The current study adapted the criteria related to the presentation of vocabulary from the checklists of AbdelWahab (2013), Mukundan et al. (2011), Miekley (2005), and Daoud and Celce-Murcia (1979). Some items were deleted because of their unsuitability. Others were divided into more specific items and a few new items were added. Ultimately, 12 items were adapted. The rating system is based on a 5-point scale: Excellent (4), Good (3), Satisfactory (2), Poor (1), and Totally lacking (0).

4.2   Phase Two

In the second phase, a semi-structured interview with the teachers was conducted. The teachers were given the prototype checklist (Appendix B). They were informed that they could remove, add, and/or revise any of the items to improve the comprehensiveness, clarity and relevance of the checklist. Then, they examined the instrument and were requested to provide their feedback in the 'comments' column next to each item. As a result of the teachers' comments, some of the items were modified, as follows:

a) **deleted item**
   1. *Good distribution of vocabulary load across the chapters* (item 3), was deleted as it was considered similar with item 1.

b) **modified items**
   1. *There is distribution of simple to complex words* (item 2), was modified to *There is balance of simple and complex words.*
   2. *Words are efficiently repeated and recycled (also use in writing and speaking for delayed recall) across the book* (item 5), was modified to *New words are repeated across the book for delayed recall.*
   3. *Words are contextualized* (item 6), was modified to *New words are contextualized.*
   4. *The sentences and examples use words that are known by learners* (item 11), was modified to *The sentences and examples that define new vocabulary use words that are known by learners.*

c) **revised item**
  1. *New vocabulary is presented in a variety of ways* (item 6) was unclear and needed to be operationalized. As one of the teachers pointed out, this item was rather ambiguous; "*in a variety of what ways*"? One of the most common ways to present new vocabulary for young learners is through visuals. Therefore, the item was revised as *Visuals have been used effectively to present the new vocabulary.*

4.3   Phase Three

The final version of the checklist (Appendix C) was developed based on the analysis of the teachers' comments as indicated by the track-changes in Appendix B. The items of the current checklist for the presentation of new vocabulary criteria are 11. Because the checklist was developed relying on its prospective users' views, it can be claimed to be easier to understand, and to be more focused on the needs and wants of the learners and educational institute. The checklist is also economical and permits a good deal of information to be recorded in a relatively short space of time.

To help evaluators interpret their assigned mean scores, the researchers added a Scores Interpretation Guide below the checklist. In order to evaluate a textbook, having read every item, evaluators assign a value of 0 to 4 for each, add up these values in order to compute the total score. This sum score is then divided by the total number of the items (i.e., 11) to calculate the mean score. Based on the Scores Interpretation Guide, if a textbook is rated with mean scores ranging from 0 to 0.80, it is categorized as being 'negligibly useful' for learners. This is followed by mean scores ranging from 0.81 to 1.60, representing 'low usefulness', from 1.61 to 2.80, showing 'moderate usefulness', from 2.81 to 3.60, indicating 'high usefulness', and finally from 3.61 to 4.00 meaning a 'very high level of usefulness'. These categories were determined according to Guilford's (1973) rule of thumb, based on which, values <20% are regarded as 'negligible', 20%-40% as 'low', 40%-70% as 'moderate', 70%-90% as 'high', and finally >90% as 'very high'.

The teachers were asked to use the checklist to evaluate the presentation of new vocabulary items in Malaysian Secondary English language textbooks. Based on their evaluation, SPSS (Version 16) was used to test the internal reliability of the instrument. Based on the reliability results, Cronbach's Alpha for the 11 items was 0.945, suggesting very high overall internal reliability of the instrument. Apart from this, the reliability of each item was also tested. The results showed that all the items were reliable (Table 1).

Table 1. Internal reliability results of each item.

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
item1	21.8000	52.200	.944	.932
item2	21.4000	59.800	.752	.944
item3	22.0000	58.500	.469	.950
item4	22.2000	51.200	.908	.933
item5	21.6000	57.300	.917	.939
item6	21.6000	61.300	.420	.950
item7	22.4000	53.300	.917	.934
item8	22.0000	56.500	.636	.944
item9	22.8000	46.200	.943	.933
item10	22.0000	54.500	.810	.938
item11	22.2000	45.200	.947	.934

Additionally, based on the item-total statistics, it could be observed that by removing items 3 and 6 the overall internal reliability would rise to 0.95. This finding has useful implications for further refinement of the instrument. The checklist was developed to evaluate the new vocabulary items in English textbook. The researchers took the elements from the vocabulary category of the previous researchers' checklists and then adapted a new checklist for the new vocabulary items in English language textbooks. The checklist provided a score for the evaluated textbook. A 'Comments' column was also added for evaluators' notes about the quality of presentation of vocabulary.

Based on the internal reliability test, all items were found to be internally consistent. The final checklist includes the following eleven items:

1. The load (number of new words in each lesson) is appropriate to the linguistic level of students.
2. There is balance of simple and complex words.
3. There is a good distribution of vocabulary load across whole book.
4. New words are repeated across the book for delayed recall.
5. New words are contextualized.
6. Visuals have been used effectively to present the new vocabulary
7. The topical nature of the vocabulary exercises is often meaningful to the students.
8. New lexical items appear in each unit.
9. There is specific method to teach new vocabulary.
10. The sentences and examples that define new vocabulary use words that are known by learners.

There is an index of new vocabulary at the end of the textbook.

DISCUSSION AND CONCLUSION

The purpose of the study was to develop a checklist that is appropriate to the present setting. The checklist was developed in three phases: developing the prototype based on reviewing the related literature and similar previous studies, semi-structured interviews with a group of English teachers helped the researchers construct its final version (Appendix C). As it was noted, the study was designed in the context of the Malaysian education context and the checklist was developed considering the views of Malaysian English language teachers. According to Mukundan et al. (2011), "experts' views on the construct and wording of the items would enable the researchers to realize the certain crucial issues that had been neglected in the development of the checklist" (p. 101). Years of experience give teachers the ability to differentiate successful from unsuccessful learning-teaching material in their present setting. In this study, the teachers were given freedom to voice their own views for each of the items in the 'comments' column of the checklist. The teachers were also free to add new items to the list if they considered them necessary. These measures were taken to ensure that the final checklist would be teacher-friendly.

Because the teachers were actively engaged in the construction process and their views were considered in developing the checklist, its items are expected to be easily comprehensible for its users. Considering the views of the prospective users of a new instrument helps developers improve the clarity and comprehensiveness of its items (Mukundan et al., 2011) contributing to the usefulness of the developed instrument (Nimehchisalem & Mukundan, 2013).

The checklist is comprehensive because it includes the important features related to the presentation of new vocabulary items in ELT textbooks. Previous checklists have not been developed specifically for evaluating the presentation of new vocabulary items in ELT textbooks. For instance, Miekley's (2005) checklist combines both vocabulary and grammar under one section. Item 1 in his checklist covers grammar while the other four items focus on vocabulary. This makes it impossible for the user to evaluate the two features separately. Additionally, some of the vocabulary items of previous checklists lack relevance. For instance, *Do the sentences and paragraphs follow one another in a logical sequence?* in the vocabulary section of Doud and Celce-Murcia (1979) was not relevant to vocabulary. For vocabulary items of AbdelWahab (2013), *The content involves culture-specific items* also was not relevant because the item measured content. These threats to construct validity were avoided in the

development of the present checklist.

Further studies can be done to improve the checklist. Surveys, field tests and validations of expert panels will certainly improve its quality. Further studies need to be carried out to refine the checklist. The findings of this study can be used as a guide to other researchers in developing similar instruments for their learning-teaching settings. The reliance of this study on the previous related literature and studies makes it a reliable and useful reference for future research in the area of ELT material evaluation. The output of this research provides a useful instrument for researchers in the area of textbook evaluation. It is also expected to help language teachers because the views of the teachers who may ultimately use it have been considered in the development of the checklist.

REFERENCES

AbdelWahab, M. M. (2013). Developing an English language textbook evaluative checklist. *IOSR Journal of Research & Method in Education*, *1*, 55-70.

Ansari, H., & Babaii, E. (2002). Universal characteristics of EFL/ESL textbooks: A step towards systematic textbook evaluation. *The Internet TESL Journal, 8*(2).

Arksey, H., & Knight, P. (1999). *Interviewing for social scientists: An introductory resource with examples.* London: Sage.

Ary, D., Jacobs, L. C., & Razavieh, A. (2002). *Introduction to research in education* (6th ed.). Belmont, CA: Wadsworth.

Breen, M., & Candlin, C. (1987). What materials? A consumer's and designer's guide. In L. E. Sheldon (Ed.), *ELT textbook and materials: Problems in evaluation and development.* ELT Documents 126 (pp. 13-28). Oxford: Modern English Publications in association with The British Council.

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

Bruder, M. N. (1978). Evaluation of foreign language textbooks: A simplified procedure. In H. S. Madsen & J. D. Bowen (Eds.), *Adaptation in language teaching.* Appendix 2 (pp. 209-218). Rowley, Mass: Newburry.

Byrd, P. (2001). Textbooks: Evaluation and selection and analysis for implementation. In Celce-Murcia, M. (Ed.), *Teaching English as a second or foreign language (3rd ed.).* Boston: Heinle & Heinle.

Canado, M. L. P., & Esteben, A. (2005). Authenticity in the teaching of ESP: An evaluation proposal. *Scripta Manent 1*(1), 35-43. Retrieved from http://www.sdutsj.edus.si/ScriptaManent/2005_1/Perez_Almagro.pdf

Cunningsworth, A. (1984). *Evaluating and selecting EFL teaching materials.* London: Heinemann Educational Books.

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

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

Daoud, A., & Celce-Murcia, M. (1979). Selecting and evaluating a textbook. In M. Celce-Murcia & L. McIntosh (Eds.), *Teaching English as a second or foreign language* (pp. 302-307). Cambridge, MA: Newbury House Publishers.

Dougill, J. (1987). Not so obvious. In L. E. Sheldon (Ed.), *ELT textbook and materials: Problems in evaluation and development.* ELT documents 126 (pp. 29-36). Oxford: Modern English Publications in association with The British Council.

Driss, L. (2006). *Systematic textbooks supplementation (knowing when to supplement and what to add).* US Department of State, Ministry of Education Morocco, University of Delaware, English Language Institute. Retrieved from http://www.udel.edu/eli/2006P4L/driss.pdf

Ellis, R. (1997). The empirical evaluation of language teaching materials. *ELT Journal*, *51*(1), 36-42.

Frazer, L., & Lawley, M. (2000). *Questionnaire design & administration.* Brisbane: John Wiley & Sons Australia, Ltd.

Garinger, D. (2001). Textbook evaluation. *TEFL Web Journal.* Retrieved from http://www.teflweb-j.org/v1n1/garinger.html

Garinger, D. (2002). Textbook selection for the ESL classroom. *Center for Applied Linguistics Digest.* Retrieved from http://www.cal.org/resources/Digest/0210garinger.html

Grant, N. (1987). *Making the most of your textbook.* London: Longman Inc.

Gray, D. E. (2009). *Doing research in the real world (2nd ed.).* London: Sage.

Griffiths, C. (1995). Evaluating materials for teaching English to adult speakers of other languages. *Forum, 33*(3), 50. Retrieved from http://forum.state.gov/vols/vol33/no3/p50.thm

Harmer, J. (1991). *The practice of English language teaching.* Harlow, Essex: Longman.

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

Haycraft, J. (1978). *An introduction to English language teaching.* London: Longman.

Hemsley, M. (1997). The evaluation of teachers' guides-design and application. *ELTED, 3*(1), 72-83.

Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centered approach.* Cambridge: Cambridge University Press.

Krug, K. (2002). *Some points to ponder in textbook evaluation.* Retrieved from http://www.kutztown.edu/library/materials/textbookeval.htm

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

Litz, D. R. A. (2005). Textbook evaluation and ELT management: A South Korean case study. *Asian EFL Journal.* Retrieved from http://www.asian-efl-journal.com/Litz_thesis.pdf

Matthews, A. (1985). Choosing the best available textbook. In A. Matthews, M. Spratt, and L. Dangerfield (Eds.), *At the chalkface* (pp. 202-206). London: Edward Arnold.

McDonough, J., & Shaw, C. (2003). *Materials and methods in ELT: A teacher's guide* (2nd ed.). Malden, MA: Blackwell Publishing Ltd.

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

Miekley, J. (2005). ESL textbook evaluation checklist. *The Reading Matrix, 5*(2).

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., 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., & Mukundan, J. (2013). Usefulness of the English Language Teaching Textbook Evaluation Checklist. *Pertanika Journal of Social Sciences and Humanities (JSSH) 21*(2), 797-816.

Peacock, M. (1997). *Choosing the right book for your class.* Retrieved from http://www.sx.ac.uk/linguistics/pgr/egspll/volume1/PDFs/PEACOCK1.pdf

Rahimy, R. (2007). A coursebook evaluation. *ESP World, 6,* 2(15). Retrieved from http://www.esp-world.info/contents.htm

Richards, J. C. (2001). *Curriculum development in language teaching.* Cambridge: Cambridge University Press.

Roberts, J. T. (1996). Demystifying materials evaluation. *System, 24*(3), 375-389.

Robinett, W. (1978). *Teaching English to speakers of other languages: Substance and technique.* Minneapolis: University of Minnesota Press.

Rubdy, R. (2003). Selection of materials. In B. Tomlinson, *Developing materials for language teaching* (pp. 37-57). London: Continuum.

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). Boston, MA: Heinlein & Heinlein publishers.

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

Tucker, C. A. (1975). *Evaluating beginning textbooks. English teaching Forum, 13(3), 355-361.*

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

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

Yap, H. P. (2011). *The evaluation of the 1st year Malaysian English primary school textbook.* Unpublished master's thesis, Universiti Putra Malaysia, Serdang.

Zabawa, J. (2001). Criteria for FCE textbook evaluation: An attempt at questionnaire formulation. Retrieved from: http://www.univ.rzeszow.pl/fil_ang/wsar2/sar_v2_17.pdf

Appendix A. Evaluative Criteria of the Checklist and their Sources (Phase 1)

Source	Item as it appeared in the Source	Item in the prototype checklist
Daoud & Celce Murcia (1979)	Does the vocabulary load (i.e., the number of new words introduced every lesson) seem reasonable for the students of that level 3?	1. The load (number of new words in each lesson) is appropriate to the linguistic level of students.
	Is the new vocabulary repeated in subsequent lessons for reinforcement?	6. Words are efficiently repeated and recycled (also use in writing and speaking for delayed recall) across the book.
	Are the vocabulary items controlled to ensure Systematic gradation from simple to complex?	2. There is distribution of simple and complex words.
Miekley (2005)	Are the new vocabulary words presented in a variety of ways (e.g. glosses, multi-glosses, appositives)?	7. New vocabulary words are presented in a variety of ways.
	Are the new words repeated in subsequent lessons to reinforce their meaning and use?	5. Words are efficiently repeated and recycled (also use in writing and speaking for delayed recall) across the book.
Mukundan et al. (2011)	The load (number of new words in each lesson) is appropriate to the level.	1. The load (number of new words in each lesson) is appropriate to the linguistic level of students.
	There is a good distribution (simple to complex) of vocabulary load across chapters and the whole book.	2. There is distribution of simple and complex words.
	Words are efficiently repeated and recycled across the book.	5. Words are efficiently repeated and recycled (also use in writing and speaking for delayed recall) across the book.
	Words are contextualized.	6. Words are contextualized.
AbdelWahab (2013)	The load (number of new words in each lesson) is appropriate to the linguistic level of students.	1. The load (number of new words in each lesson) is appropriate to the linguistic level of students.
	There is a good distribution (simple to complex) of vocabulary load across the whole book.	2. There is distribution of simple and complex words. 3. Good distribution of vocabulary load across the chapters. 4. Good distribution of vocabulary load across the whole book.
	Words are contextualized.	6. Words are contextualized.
	The topical nature of the vocabulary exercises is often meaningful to the students.	8. The topical nature of the vocabulary exercises is often meaningful to the students.
AbdelWahab (2013)	New lexical items appear in each unit.	9. New lexical items appear in each unit.
	There is specific method to teach new vocabulary.	10. There is specific method to teach new vocabulary.
	The sentences and examples use words that are known by learners.	11. The sentences and examples use words that are known by learners.
	There is a list of vocabulary items tagged at the end of the textbook.	12. There is an index of new vocabulary at the end of the textbook.

Appendix B. Prototype Checklist for the Items of New Vocabulary Refined (Phase 2)

Note: The following track changes indicate the changes made to the checklist after the second phase.

~~Vocabulary~~Item*	0	1	2	3	4	Comment
1. The load (number of new words in each lesson) is appropriate to the linguistic level of students.						
2. There is ~~distribution~~ balance of simple and complex words.						
~~Good distribution of vocabulary load across the chapters.~~						*Teacher comment: This is something like item 1. Remove it!*
3. Good distribution of vocabulary load across the whole book.						
4. N~~ew~~ ~~Words~~ words are ~~efficiently~~ repeated ~~and recycled (also use in writing and speaking for delayed recall)~~ across the book for delayed recall.						
5. N~~ew~~ ~~Words~~ words are contextualized.						
6. ~~New words are presented in a variety of ways.~~ Visuals have been used effectively to present the new vocabulary						*Teacher comment: Item 6 is unclear and need be operationalized. This item is rather ambiguous; "in a variety of what ways"?*
7. The topical nature of the vocabulary exercises is often meaningful to the students.						
8. New lexical items appear in each unit.						
9. There is specific method to teach new vocabulary.						
10. The sentences and examples that define new vocabulary use words that are known by learners.						
11. There is an index of new vocabulary at the end of the textbook.						

Appendix C. Checklist for Evaluating New Vocabulary Presentation in ELT Textbooks (Final Version)

Introduction:
This checklist has been developed to help you evaluate the presentation of new vocabulary items in Malaysian English Language Textbooks.

Background information

1. Age: …… years
2. Level of education: ☐ Diploma   ☐ BA   ☐ MA   ☐ PhD
3. Teaching experience ….. years
4. Textbook title: ……………………………………

Instructions:
Read the items below carefully and mark the appropriate number that best describes your evaluation of the textbook:

0: Totally lacking   1: Poor   2: Satisfactory   3: Good4: Excellent

If you have any further comments about each item, you may leave your notes in the '*Comments*' column. After calculating the mean score, you can interpret it using the 'Scores Interpretation Guide' below the checklist.

#	Item	0	1	2	3	4	Comments
1.	The load (number of new words in each lesson) is appropriate to the linguistic level of students.						
2.	There is balance of simple and complex words.						
3.	Good distribution of vocabulary load across the whole book.						
4.	New words are repeated across the book for delayed recall.						
5.	New words are contextualized.						
6.	Visuals have been used effectively to present the new vocabulary.						
7.	The topical nature of the vocabulary exercises is often meaningful to the students.						
8.	New lexical items appear in each unit.						
9.	There is specific method to teach new vocabulary.						
10.	The sentences and examples that define new vocabulary use words that are known by learners.						
11.	There is an index of new vocabulary at the end of the textbook.						

Scores Interpretation Guide

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