

# The Retrospective Evaluation of English Language Teaching Materials in a Preparatory Year English Language Program in Saudi Arabia

Abdulaziz Alfehaid<sup>a</sup>, Nada Alkhatib<sup>a</sup>, Jayakaran Mukundan<sup>\*b</sup>

<sup>a</sup>Imam Abdulrahman Bin Faisal University, Saudi Arabia

<sup>b</sup>Taylors University, Malaysia

*\*Corresponding author: jayakaran.mukundan@taylors.edu.my*

Article information	
<b>Abstract</b>	<p>The purpose of this study was to retrospectively evaluate the use of English language teaching materials in a one-year preparatory program of a Saudi Arabian public university. The materials were a set of ten books (commercial coursebooks) by Oxford University Press, consisting of five-level listening and speaking (five books) and reading and writing books (five books), with teacher-produced supplementary material (five books). Ninety-six teachers took part in the first part of the study, a macro-evaluation, by use of an instrument, the ELT-TEC. Then, the materials were micro-evaluated in-use through classroom observations and interviews. The results revealed a significant difference in the performance of the two book types. The commercial coursebooks books were rated more highly than the teacher-produced textbooks in macro evaluations. The observations and interviews confirmed the strengths of the commercial coursebooks, which showed high satisfaction/suitability in all sections across the ELT-TEC, while the teacher-produced textbooks revealed weaknesses in four major sections. As this evaluation exercise was done to</p>

	determine the effectiveness of the materials used in the preparatory year, college administrators would have to make crucial decisions on the fate of the teacher-produced books as they have been found to have numerous weaknesses and decide if they have to be omitted or if an overhaul would be needed.
<b>Keywords</b>	teacher-produced textbooks, commercial coursebooks, macro- and micro-level textbook evaluation, retrospective evaluation
<b>APA citation:</b>	Alfehaid, A., Alkhatib, N., & Mukundan, J. (2024). The retrospective evaluation of English language teaching materials in a preparatory year English language program in Saudi Arabia. <i>PASAA</i> , <i>69</i> , 335-369.

## 1. Introduction

Even in the not-so-distant past (as far as three decades ago), the selection of textbooks seemed to be the only reason why teachers needed to know anything about textbook evaluation. As a result of this, predictive evaluation, or evaluation of textbooks for the purposes of selection, seemed to be the only major concern of teachers. Traditionally this form of evaluation was done with instruments such as checklists and questionnaires (Ellis, 1997). The overemphasis placed on predictive evaluation in fact fueled the proliferation of evaluation instruments developed for the sole purpose of predictive evaluation (Mukundan, 2010).

Ellis (1997, p. 37) laments the dearth of “published accounts of retrospective evaluations of course materials,” which many now believe should be considered more important than predictive evaluation. He also states that the closest to empirical evaluation of teaching materials is the one account of findings reported on the trialing of new materials (Barnard & Randall, 1995). Retrospective evaluation of textbooks has only recently been researched with enthusiasm (Ton Nu & Murray, 2020), and these new developments show promise as materials

evaluated retrospectively inform teachers of the performance of materials as well as provide opportunities for adaptation (Ton Nu & Murray, 2020).

Part of the reason why retrospective evaluation is not widespread is that teachers may not have adequate knowledge on how it is done. This may also be due to the lack of initiatives taken by institutions to make materials evaluation a part of continuous professional development. The retrospective evaluation of textbooks must be seen as a long-term investment in teacher professional development. Traditionally, teachers have been doing adaptations because they know from intuitions and impressionistic evaluations what works and what does not (Ellis, 1997). Some teachers intuitively test the materials that they use in the classroom, and their actions, observations, and judgment sometimes lead to adaptations. Unfortunately, not many of these experiences (which are intuitive by nature) are reported. Recent research on teachers' sense of plausibility (TSOP) also suggests that teachers, no matter how much knowledge they have acquired on teaching materials, abandon this acquired knowledge and trigger their TSOP, something that is intuitive and is acquired through experience (Prabhu, 2019). This allows their intuitions to take over and critically examine the materials used, sometimes omitting that which is considered not applicable in their teaching contexts.

Regarding the retrospective evaluation of materials to inform the larger community of teachers, there is a need to coerce teachers into the empirical examination of materials (Ellis, 1997; Mukundan, 2010) so that not only are the strengths and weaknesses of the materials identified, but adaptations are also put in place. The bigger advantage of the empirical retrospective evaluations of textbooks is that it promotes teachers' professional development initiatives (Mukundan, 2010). Teachers, while evaluating materials retrospectively, are in fact involved in research. A major challenge in fostering the retrospective evaluation of materials would be in the methodology and procedures that need to be

established. The predictive evaluation of materials would probably need a single instrument. While it is still possible to use an instrument that is generally used for predictive evaluation as one for retrospective evaluation, it may be too restrictive. Retrospective evaluation, thus, would need to probe the performances of materials from multiple perspectives, such as their performance in the classroom. This would need something that resembles a composite framework (Mukundan, 2010), where data are mined through multiple instruments and thereby satisfying the needs for triangulation and providing depth in the probe. The novelty within the current study is that it combines both macro and micro evaluations of textbooks, and in-use (which has never been done elsewhere before), thereby strengthening the evaluation process through the triangulation of data.

## **2. Literature on the Retrospective Evaluation of Teaching Materials**

Ellis (1997), Mukundan (2010), and Mukundan and Ahour (2010) have critically examined the types of materials evaluation activities carried out before 2000 and have suggested that most of the literature on materials evaluation has been inundated with reports on instrument development for predictive evaluation (selection purposes). However, the new millennium saw some important changes in the field of materials evaluation, where investigations of textbooks that were predominantly in-use and retrospective have gained ground. While the literature sometimes suggests that there are differences between retrospective evaluation and in-use evaluation (Lisna, 2016), it is widely believed now that in-use and retrospective evaluation can complement each other as the evaluation of books used (and that are still in use) will benefit teachers from the perspective of materials adaptation (Mukundan, 2007, 2010) and teachers' professional development (Richards & Lockhart, 1994).

Research on textbook evaluation can be done at the macro- or micro-level (Ellis, 1997). Most research in the early years of the new millennium focused on macro-level evaluations, which exclusively used evaluation checklists (Gholami et

al., 2017). Mukundan and Kalajahi (2013) conducted the first state-wide retrospective evaluation of English language textbooks in Malaysia, involving 944 teacher-participants in the state of Melaka. All of the English language books used at the primary level (Years 1-6) and secondary level (Forms 1-5) were evaluated by the teachers. The evaluation revealed that the primary books were more acceptable (highly beneficial) than the secondary books (moderately useful). While the writers focused on overall assessment, the pressing question in the minds of the readers would be the reasons why the secondary school books were only moderately useful. Surveys like this that only provide quantitative data from evaluation checklists cannot provide clear information on specific problems within the materials used.

Other smaller-scale macro-level studies were done in some other parts of the world. Kirkgoz (2011), for example, evaluated primary school English textbooks that were used in Grade 4 in Turkey, involving teachers and learners as informants, and found data from both groups showing congruence. Smaller, institutional-level studies were done by Nemati (2009) on an English pre-university textbook in Karnataka state in India and the results indicated that the book was acceptable except in the domain of vocabulary learning, where the informants claimed that the grading of the vocabulary was poor. Chao (2010), on the other hand, in his retrospective study of an English textbook used in a local college in Taiwan that focused on intercultural competence reported that the textbook “did not seem to help learners develop intercultural abilities” (p. 40).

While most retrospective evaluations of textbooks at the macro level have teachers as participants, the study by Tang and Zheng (2018) was done on students. The researchers used an adaptation of McDonough and Shaw’s (2003) two-procedure evaluation model to develop a questionnaire. This study was unique, as students were the only participants of the research. It also revealed the importance of research of this nature. Students were highly critical of the two

books, which were used widely in China. Useful feedback from students, such as “the passages were too difficult for the students to understand” and “it had too many translation exercises” (p. 60), were indicators of the factors contributing to learner fatigue, something that should be considered in future adaptations of the book.

A paradigm shift in materials evaluation practices came about when micro studies became the preferred method of investigation. Micro studies that involved content analysis as the main investigation procedure were more frequently done at the beginning of the 1990s (Boxer & Pickering, 1995; Gilmore, 2004; Littlejohn, 1998, 2022; Uso-Juan, 2007; Zarifi & Mukundan, 2012). These researchers investigated specific issues in materials, which were carried out at the micro level. Ton Nu and Murray (2020), for instance, found that Vietnamese textbooks aimed at getting learners to achieve communicative competence but failed in terms of content, which would have helped with the acquisition of pragmatic competence, while a corpus-based study by Zarifi and Mukundan (2012) looked at how the teaching of phrasal verbs in textbooks could be problematic.

The introduction of corpus-based textbook evaluation studies (Mukundan, 2010) led to the digitization of textbooks, which in turn led to numerous micro investigations (Littlejohn, 2022) of the performance of textbooks in terms of vocabulary load and distribution patterns, using concordance software. Some of these studies were forensic in nature, investigating how *dead* books could be viewed as having contributed to poor learning and the acquisition of English in countries like Malaysia, where English is a second language. To further exemplify, Mukundan and Aziz (2009) found serious irregularities in the load and distribution patterns within all the five books that were investigated. In addition, they also found that words specified in syllabus wordlists were sometimes not present in any of the five books, while words that were not important and not listed in the major wordlists such as the *Oxford 3000* were present in abundance. An even more

serious revelation was that the repetition of words within the book was minimal and the recycling of words almost negligible. Studies like this soon led to other researchers investigating the performance of textbooks from the perspective of the representations of the parts of speech within them. The study of the semantic treatment of phrasal verbs (Zarifi & Mukundan, 2015) within textbooks and textbook representation of prepositions (Mukundan & Roslim, 2009) were such studies.

The literature on the evaluation instruments used in a retrospective evaluation of ELT materials differs for macro- and micro-level evaluations. Macro-level evaluations are straightforward, requiring “an overall assessment of whether an entire set of materials has worked” (Ellis, 1997, p. 37). McGrath (2002) and Kashoob (2018) elaborated on this and stated that macro-level evaluations are inclusive compared to micro-level evaluations. They also suggest that they can be done separately or together (as in the present study). Macro-level evaluations, because they sometimes require a large sample, require the use of instruments also commonly referred to as textbook evaluation checklists. Studies like the ones by Mukundan and Kalajahi (2013), which were conducted statewide and evaluated 11 textbooks in the school system, are ones totally dependent on the checklist. Micro-level studies, however, are done in order to investigate the specific features of textbooks and instruments developed, for these vary according to the nature of the study. There are various types of approaches used in micro-level studies, and one of the most cited (Littlejohn, 2022) is that of the “three levels analysis” (Littlejohn, 1998, 2022), which moves from objective description through subjective analysis to subjective inference. Ton Nu and Murray (2020), on the other hand, in their micro-level study, devised an instrument based on a framework (which was a combination of the two others) to evaluate the pragmatic content in EFL textbooks.

The theoretical basis for this research came from the composite framework for ESL textbook evaluation (Mukundan, 2010). This framework recommends the triangulation of data from multiple sources in the retrospective evaluation of textbooks. The composite framework, which has three distinct instruments (the checklist, the teacher's log, and the concordance software), worked in unison (with interdependence on each other) in the evaluation of textbooks. This composite framework was inspired by Richards and Rogers' (1982) model for methodology, which shows the interactions among the three components: design (representing material) interacting with approach (theories of language and language learning), and procedure (activities, techniques, and practices within the learning-teaching environment). It is evident from Richards and Rogers' model that "materials do not exist in isolation" (Mukundan, 2010, p. 17), and any evaluation of materials must not treat examination in isolation, i.e., the context (especially the classroom) must always be in view. The materials (design) investigated must be viewed from multiple perspectives, especially that of the interactions that take place from the influences of approach and procedure (Richards & Rogers, 1982).

No studies have been found in the literature reporting on the retrospective evaluation of English language teaching materials utilizing both macro- and micro-level investigations. The previous studies on retrospective evaluation in the past two decades have either been at the macro level, such as a study by Mukundan and Kalajahi (2013), which was an evaluation of the textbooks used in an entire state, or the more popular micro investigations such as the one done by Ton Nu and Murray (2020). The present study was more comprehensive in the sense that it had aspects of both macro and micro (Ellis, 1997) investigations, thereby fulfilling some important data triangulation requirements, which are crucial in studies of this nature. This would help fill some of the research gaps in the existing literature.

The three research questions responded to in this study are as follows:



1. What are the strengths and weaknesses evident in the macro-level investigations of the ELT materials used in a preparatory year program in Saudi Arabia?
2. To what extent do the micro-level evaluations of the materials in-use provide supporting evidence for macro-level investigations?
3. How do teachers cope with shortcomings of materials in the classroom and how does this influence their perceptions of textbooks?

### **3. Method**

The first part of the study involved the macro evaluation of the textbooks. This was done quantitatively. A textbook evaluation checklist, the ELT-TEC (Mukundan & Nimehchisalem, 2012), was used in order to elicit data. The second part of the study, micro evaluation, focused on classroom observations (where the books were used) and interviews with the teachers and learners using the textbooks. Data from this part of the study were analyzed and reported qualitatively.

#### **3.1 The Teaching Materials Under Investigation**

This study evaluated the teaching materials used in a one-year preparatory program at a university in Saudi Arabia. The materials investigated came under two categories: the first category consisted of the commercially produced coursebooks *Q: Skills for Success* (Oxford University Press), which had five levels, each of which was dedicated to a listening and speaking book and a reading and writing one. There was a total of ten books in this category, which will henceforth be referred to as the commercial coursebooks. The second category consisted of the teacher-produced textbooks (four books for the four different tracks—health, science, engineering, and humanities), and another book used by those in the health, science, and engineering tracks. These textbooks will be referred to as the teacher-produced textbooks, as several teachers were assigned to develop them.

### **3.2 Instruments**

For the first part of the study, which involved the macro-level evaluation of the textbooks, the instrument used was the ELT-TEC-English Language Teaching-Textbook Evaluation Checklist (Mukundan & Nimehchisalem, 2012). It has two main sections: section A: general attributes and section B: learning-teaching content. General attributes considers the following aspects: a) the book in relation to the syllabus and curriculum (in figures and tables referred to as SC-fit), b) methodology, c) suitability for learners, and d) supplementary materials. The aspects under section B: learning-teaching content, evaluates a) general aspects, b) listening, c) speaking, d) reading, e) writing, f) vocabulary, g) grammar, h) pronunciation, and i) exercises.

As the entire materials package consisted of the commercial coursebooks and the teacher-produced textbooks, two experts on materials evaluation, both professors, were consulted in order to determine if the ELT-TEC was suitable for evaluating the different types of textbooks. They stated that the instrument was suitable for any English language textbook and recommended that all of the books be evaluated using the same instrument. After analysis of both types of textbooks, the two expert validators concluded that while the teacher-produced textbooks were originally produced to immerse preparatory program students into ESP and EAP, there was no evidence to support the claim that they were ESP (Hutchinson & Waters, 1987) or EAP (Jordan, 1997). All of the textbooks resembled regular textbooks.

The ELT-TEC instrument has been widely considered user-friendly, requiring informants to complete the evaluation in only 9.5 minutes (Mukundan & Nimehchisalem, 2012). When tested for concurrent validity with other well-established instruments like that of Skierso (1991), strong correlation was achieved ( $r = 0.75$ ). Inter-rater reliability tests showed acceptable reliability coefficients ( $r = 0.962$ ). There were, however, in the present study, minor

adaptations done to the ELT-TEC, where six items related to the physical and utilitarian aspects of the book were removed based on the recommendations of administrators of the preparatory program. After the adaptation, another reliability analysis was carried out, and it showed an internal reliability consistency score (Cronbach's alpha) of 0.98. This was above the recommended reliability cut-off value of 0.7 (Field, 2013). The second instrument was the structured-interview protocol, which was developed for teachers and students.

### **3.3 The Participants**

There were 96 registered teachers teaching English in the preparatory year. Those without administrative duties taught an average of 20 hours per week. Nineteen teachers were randomly selected to be observed in classrooms and to be interviewed. Seven taught using the teacher-produced textbooks, while the rest (12) taught using the commercial coursebooks. In addition, 30 students (ten from each of the three levels: elementary, intermediate, and advanced) were interviewed. Ethical clearance for this research was obtained from the ethics committee of the university. Participation was voluntary. The participants were briefed on the research and were told that their names and other details would not be revealed. All of the participants signed the informed consent form before their involvement in the research.

### **3.4 Data Collection**

The 96 registered teachers contributed to a total of 193 responses (the teachers taught different classes, sometimes using two or three books) to the ELT-TEC—142 responses to the survey on the commercial coursebooks and 51 to the survey on the teacher-produced textbooks. The average teaching experience was recorded at 8.84 years ( $SD = 2.7$ ).

A total of 19 teachers (three females and 16 males) were observed in the classrooms (the male teachers taught the male students, and the female teachers

taught the female students). Each was observed for an hour, and a total of 19 hours of classroom observations were carried out. As Covid-19 was still an issue at the time of the observations, classes of 60 were reduced to half (face-to-face), while the rest participated through Zoom (they took turns doing this). One of the researchers (a non-staff member, a visiting research professor) observed all of the classes taking field notes (excerpts and references to observation field notes were labelled OFN). At the end of each class, teacher-researcher conferences took place. Interviews were also arranged on separate days in order to gather more data. All 19 teachers were interviewed for about 20 minutes (a total of 380 minutes, approximately six hours). The strict conservative values emphasized there did not allow for female students to be exposed to males (teachers and researchers). The researcher that did the observations was male, so the female classes were observed online (with the camera shut off). Interviews with female teachers was however possible face to face. Interviews with female students were not possible. Thirty students (only males allowed) were interviewed (each for five minutes, for a total of 150 minutes).

In order to reduce researcher bias, the interview protocol, classroom observation write-up templates, and data collection and analysis procedures were validated by two professors, both experts in materials evaluation. The data collected from the participants also underwent member checks. The researchers were also constantly aware of the threat of participant bias, so responses that were suspicious or doubtful were rephrased or asked in different ways.

### **3.5 Data Handling and Exclusion Criteria**

Prior to the data analysis, the data were explored, and all of the participants using a single-response type in the survey were removed. This resulted in the removal of ten participants from the data, i.e., four from the survey on commercial coursebooks and six from the survey on teacher-produced textbooks. Therefore, a total of 183 responses from the participants was used in the subsequent analyses.

### 3.6 Data Analysis

The data collected through the ELT-TEC (for RQ 1) were analyzed using IBM SPSS statistics, while those of the observations and interviews were done through deductive thematic analysis based on the criteria in the ELT-TEC. In order to interpret the mean scores, the scores interpretation guide (Nimehchisalem & Mukundan, 2015, p. 780) was used. According to this guide, the mean values are categorized under five levels as follows:

< 0.80: Negligible usefulness

0.81-1.60: Low usefulness

1.61-2.80: Moderate usefulness

2.81-3.60: High usefulness

> 3.61: Very high usefulness

## 4. Findings and Discussion

The researchers began this section by reporting the results of the first research question, which was as follows:

RQ 1: What are the strengths and weaknesses evident in the macro-level investigations of the ELT materials used in a preparatory year program in Saudi Arabia?

The results are reported graphically in Table 1 and Figure 1 below.

**Table 1**

*Teachers' Rating of Commercial Coursebooks and Teacher-produced Textbooks*

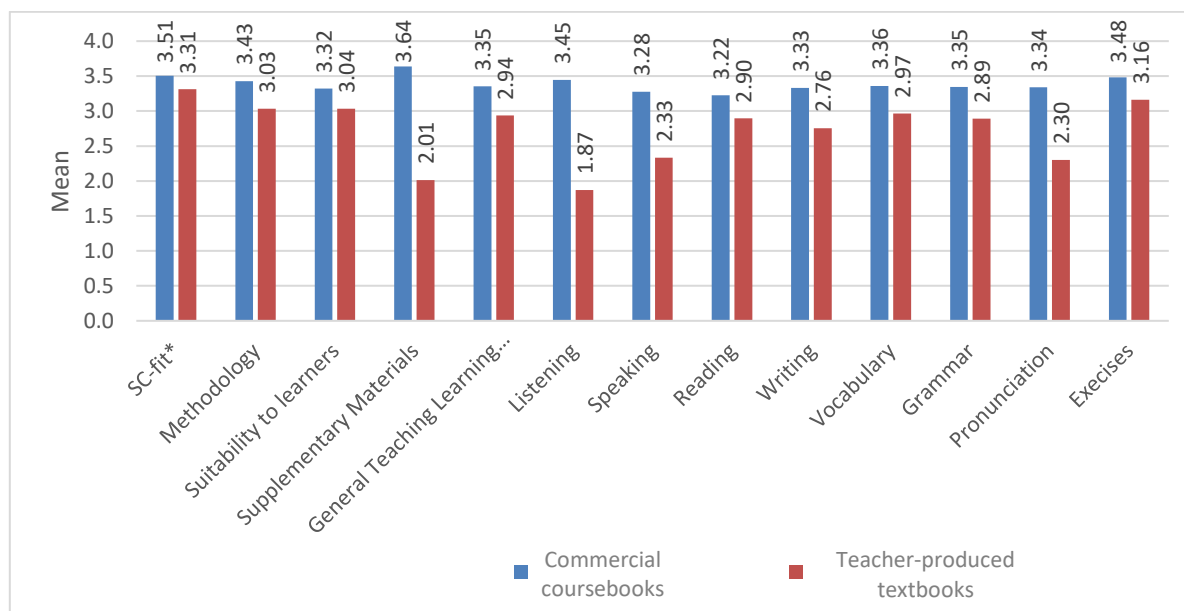
	Commercial coursebooks			Teacher-produced textbooks		
	Mean	SD	Satisfaction	Mean	SD	Satisfaction
SC-fit*	3.51	.58	High	3.31	.85	High
Methodology	3.43	.70	High	3.03	.98	High
Suitability	3.32	.69	High	3.03	.91	High
Supplementary Materials	3.64	.55	Very high	2.01	1.37	Moderate

	Commercial coursebooks			Teacher-produced textbooks		
	Mean	SD	Satisfaction	Mean	SD	Satisfaction
General Learning- Teaching Content	3.35	.65	High	2.93	.87	High
Listening	3.44	.63	High	1.87	1.47	Moderate
Speaking	3.28	.76	High	2.33	1.38	Moderate
Reading	3.22	.94	High	2.90	1.04	High
Writing	3.33	.74	High	2.75	1.13	Moderate
Vocabulary	3.36	.77	High	2.97	.92	High
Grammar	3.35	.72	High	2.89	1.02	High
Pronunciation	3.34	.79	High	2.30	1.61	Moderate
Exercises	3.48	.65	High	3.16	.95	High

\* Note: SC-fit stands for the “Book in Relation to Syllabus and Curriculum.”

**Figure 1**

*Teachers’ Ratings of Commercial Coursebooks and Teacher-produced Textbooks*



The results in Figure 1 show that the participants’ overall satisfaction ratings of the commercial coursebooks were all higher than those of the teacher-produced textbooks. It is obvious that the satisfaction ratings were significantly different in five book-specific attributes, namely supplementary materials, E-ISSN: 2287-0024

listening, speaking, writing, and pronunciation. To further explain, the commercial coursebooks recorded very high or high levels of satisfaction relative to the teacher-produced textbooks that recorded only high or moderate levels of satisfaction (see Table 1 above).

However, to further explore the participants' ratings of each attribute of the corresponding materials, only the learning-teaching content was considered at the next stage due to its relevance to the micro-level investigations, which investigated what went on in the classroom.

#### 4.1 Macro Evaluation of the Commercial Coursebooks and Teacher-Produced Textbooks in Relation to the Learning-Teaching Content

The study explored the participants' satisfaction/suitability ratings of the learning and teaching content of the commercial coursebooks and teacher-produced textbooks. This covered general content, listening, speaking, reading, writing, vocabulary, grammar, pronunciation, and exercises. The results of the analysis are shown in Table 2 below.

**Table 2**

*Teachers' Ratings of Learning-Teaching Content of the Commercial Coursebooks and Teacher-Produced Textbooks*

Item	Commercial coursebooks			Teacher-produced textbooks		
	Mean	SD	Satisfaction/ Suitability	Mean	SD	Satisfaction/ Suitability
1. Most of the tasks in the book are interesting.	3.20	.89	High	2.76	1.17	Moderate
2. Tasks move from simple to complex.	3.38	.72	High	2.96	.99	High
3. Task objectives are achievable.	3.44	.70	High	2.98	.94	High
4. Cultural sensitivities have been considered.	3.54	.64	High	3.38	.84	High

Item	Commercial coursebooks			Teacher-produced textbooks			
	Mean	SD	Satisfaction/ Suitability	Mean	SD	Satisfaction/ Suitability	
5. The language in the textbook is natural and real.	3.44	.73	High	3.07	1.14	High	
6. The situations created in the dialogues sound natural and real.	3.36	.80	High	2.98	1.21	High	
7. The material is up to date.	3.21	.87	High	3.07	1.01	High	
8. It covers a variety of topics from different fields.	3.51	.67	High	3.11	1.05	High	
9. The book contains fun elements.	3.08	1.0	High	2.13	1.50	Moderate	
10. The book has appropriate listening tasks with well-defined goals.	3.38	.74	High	1.47	1.62	Low	
Listening	11. Instructions are clear.	3.64	.60	Very high	2.13	1.60	Moderate
	12. Tasks are efficiently graded according to complexity.	3.40	.76	High	1.89	1.54	Moderate
	13. Tasks are authentic or close to real language situations.	3.36	.76	High	2.00	1.61	Moderate
Speaking	14. Activities are developed to initiate meaningful communication.	3.29	.80	High	2.29	1.44	Moderate
	15. Activities are balanced among individual response, pair work, and group work.	3.41	.74	High	2.36	1.45	Moderate
	16. Activities motivate students to talk.	3.14	.96	High	2.36	1.38	Moderate
Reading	17. Texts are graded.	3.35	.94	High	2.89	1.21	High
	18. Length is appropriate.	3.20	1.00	High	2.93	1.05	High



	Item	Commercial coursebooks			Teacher-produced textbooks		
		Mean	SD	Satisfaction/ Suitability	Mean	SD	Satisfaction/ Suitability
	19. Texts are interesting.	3.12	1.06	High	2.87	1.12	High
Writing	20. Tasks have achievable goals and take into consideration learner capabilities.	3.34	.77	High	2.80	1.19	Moderate
	21. Models are provided for different genres.	3.48	.75	High	2.73	1.25	Moderate
	22. Tasks are interesting.	3.18	.90	High	2.73	1.19	Moderate
Vocabulary	23. The load (number of new words in each lesson) is appropriate to the level.	3.30	1.01	High	3.02	1.08	High
	24. There is a good distribution (simple to complex) of vocabulary load across chapters and the whole book.	3.41	.76	High	2.91	1.10	High
	25. Words are efficiently repeated and recycled across the book.	3.32	.84	High	2.84	1.04	High
	26. Words are contextualized.	3.40	.81	High	3.09	0.92	High
Grammar	27. The spread of grammar is achievable.	3.50	.67	High	2.98	1.05	High
	28. The grammar is contextualized.	3.33	.75	High	2.93	1.07	High
	29. Examples are interesting.	3.20	.95	High	2.76	1.17	Moderate
	30. Grammar is introduced explicitly.	3.37	.82	High	3.00	1.07	High
	31. Grammar is reworked implicitly throughout the book.	3.33	.91	High	2.80	1.09	Moderate
Pronu	32. It is contextualized.	3.37	.76	High	2.22	1.70	Moderate

	Item	Commercial coursebooks			Teacher-produced textbooks		
		Mean	SD	Satisfaction/ Suitability	Mean	SD	Satisfaction/ Suitability
	33. It is easy to learn.	3.31	.88	High	2.38	1.63	Moderate
Exercises	34. They have clear instructions.	3.60	.63	High	3.33	0.95	High
	35. They are adequate.	3.52	.71	High	3.11	0.98	High
	36. They help students who are under/over-achievers.	3.33	.89	High	3.04	1.15	High

As evident in Table 2, the participants seemed to be largely satisfied with the commercial coursebooks' learning-teaching content. The participants on average recorded high levels of satisfaction with the books' general content, listening, speaking, reading, writing, vocabulary, grammar, pronunciation, and exercises. In the listening section, item 11, "Instructions are clear," was rated very high. Evaluations done elsewhere on the use of the commercial coursebooks in university settings showed similar results. The study by Ahour and Ahmadi (2012), a retrospective evaluation of a textbook used in an Iranian university, revealed that the book scored highly on subject matter, vocabulary and structure, and exercises. The interviews conducted on instructors further confirmed high ratings in four components. Similar findings were reported by Lisna (2016) in her textbook evaluation study, with all areas scoring 70% or more.

The teacher-produced textbooks did not perform as well as the commercial coursebooks (Table 2) in aspects associated with the learning-teaching content. Under general features, the books performed at a high level for items 2-8 but moderately (for item 1). For the listening section, the books were rated moderately (for items 11-13) and low for item 10. For speaking, all of items (items 14-16) were rated moderately. Reading was rated high in all items (items 17-19), while for writing, items 20-22 were rated as moderate. moderately. Vocabulary, like reading,

was rated high for all items (items 23-26), grammar (items 27, 28, and 30) was rated high, while items 29 and 31 were rated moderately. Pronunciation was rated moderately for both items (items 32 and 33), and exercises were rated high for all items (items 34-36).

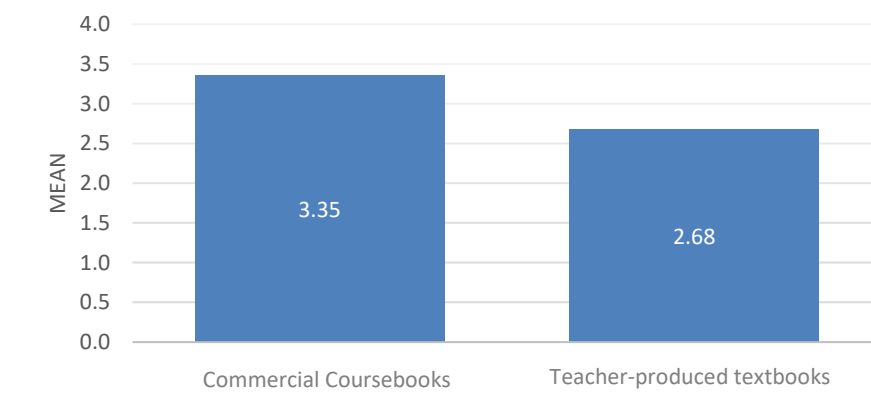
Relatively low ratings for the teacher-produced textbooks were also evident in Maryam and Sara's (2014) study, with the book in her study scoring poorly in most categories. The poor ratings for the teacher-produced textbooks, especially those produced in-house as in the study of Davari et al. (2013), also seemed to mirror the poor ratings of similarly produced books evaluated in this study. The study by Ebadi and Farjad (2015), in which they evaluated two EAP books for medical students, one produced in-house and the other by a reputable publisher, revealed vast differences between the two. The in-house publication was rated poorly by the students and was considered not appealing. A lot of problems that are linked to teacher-produced textbooks also seem to stem from "teachers' lack of expertise in producing tailor-made materials" (Littlejohn, 2022, p. 330), as they are found to lack subject-specific and pedagogic knowledge.

#### **4.2 Macro-Level Review of the Learning and Teaching Content Across the Commercial Coursebooks and Teacher-Produced Textbooks**

The study aimed to compare the learning and teaching content of the two types of books in order to find out the differences in the teachers' satisfaction. The researchers first summed up the mean score of the teachers' ratings of the general content attributes, listening, speaking, reading, writing, vocabulary, grammar, pronunciation, and exercises of the two types of books. The mean score of the overall ratings is presented in Figure 2 below.

## Figure 2

*Overall Ratings of the Learning-Teaching Content in the Commercial Coursebooks and Teacher-Produced Textbooks*



Subsequent to this, the Mann-Whitney U test was run in order to find if the teachers' ratings were significantly different across the two book types. Upon checking the distribution of normality, it was found that the data had violated the assumptions of normality. The results revealed a significant difference between the teachers' overall ratings of the commercial coursebooks ( $Mdn = 3.41$ ,  $n = 138$ ) and teacher-produced textbooks ( $Mdn = 2.74$ ,  $n = 45$ ), ( $U = 1747$ ,  $Z = -4.432$ ,  $p < .001$ ). To further explore the differences in the teachers' ratings of each specific attribute of the corresponding book in their learning and teaching content, the Mann-Whitney U test was conducted between the two books in terms of listening, speaking, reading, writing, vocabulary, grammar, pronunciation, and exercises, and significant differences in the teachers' ratings of these learning and teaching sections were found (all  $p$ -values  $< .01$ ), except for exercises ( $p = .064$ ).

### 4.3 Micro Evaluations of Materials

This section addresses the second research question:

RQ 2: To what extent do the micro evaluations of the materials in-use provide supporting evidence for macro-level investigations?

The following abbreviations are used in the presentation of the qualitative findings: observation field notes (OFN), teacher interviews (TI), and student interviews (SI).

#### **4.4 General Overview**

The data from the micro evaluations mostly confirmed the quantitative analysis of the data from the ELT-TEC in the previous discussion. All 19 teachers claimed that they were “very comfortable” using the commercial coursebooks, stating that the commercial coursebooks were “more organized” and “pedagogically sound.” When asked to elaborate on these statements, all 19 teachers (all of them teaching using both commercial coursebooks and teacher-produced textbooks) stated that a “comprehensive material outlay with supportive online learning systems which incorporate supplementary learning-teaching resources” made the commercial coursebooks “a teacher’s joy.” This supported the macro-level evaluation of the commercial coursebooks, which were rated very high for supplementary materials (Table 1). The teachers, while using the e-book version of the commercial coursebooks, had at their disposal all of the audio and video materials (all professionally produced by the publisher), which they could use by “simply clicking links on the screen, and even repeating listening/viewing input was easy.” The digital tools within the coursebook, while at use in the face-to-face mode and on the on-line platform, were also used for developing the students’ receptive and productive skills, thus suggesting that these were efficiently developed blended-learning materials that were “channelled by tools such as voice recorders which can be shared by posting on on-line platforms” (Hartle, 2022, p. 400). The teacher-produced textbooks lacked the support of online learning resources. These books were rated only moderately for the supplementary materials (Table 1). Obviously, the workload of the teachers that taught an average of 20 hours had a significant bearing on the teachers’ preferences for the commercial coursebooks. The teacher-produced textbooks,

the participants claimed, required them to “prepare for lessons much more and there seemed to be a feeling of fatigue,” every time they entered the classroom.

Likewise, the data from the micro evaluations (classroom observations and interviews) of the teacher-produced textbooks confirmed the quantitative analysis of the data from the ELT-TEC. All 19 teachers confirmed that the teacher-produced textbooks were not suited to the teaching of “restless young adults” and claimed that “there was nothing fun about them.” This supported the quantitative data from the macro-level evaluations, which rated the teacher-produced textbooks only moderately for item 1 “Most of the tasks in the book are interesting” and item 9 “The book contains fun elements” (Table 2). Research in the area of materials development has shown that only materials that are interesting and that engage learners contribute to effective learning (Tomlinson, 2011; Tomlinson and Masuhara, 2021).

The teachers elaborated at length on the teacher-produced textbook for the health track. They claimed that the book had “a strange structure, with eight units in total, and all units represented the various fields within healthcare: medicine, dentistry, pharmacy, nursing, pharmacy, and applied medical sciences.” The participants believed that “the range of fields was too diverse,” while observations in the lesson (health track) in which the teacher-produced textbook was used revealed “a futile attempt at exposing students to the language of their discourse communities, as the sub-fields were too diverse.” The teachers as well as the students indicated that the topics in the book as a result of this wide diversity made the book “boring” (SI) and “inappropriate” (TI). The teachers believed that all eight units were “horribly dry,” especially ones like “Anatomical Structure of the Human Body” and “Teeth and their Functions.” The teachers also claimed that what made it worse was that “All the reading passages were too short” and “The absence of listening tasks which could have added variety to classroom procedures” made the use of the book “resemble a chore.”

Next, an in-depth analysis of the findings from the observations on one of the 12 classes where commercial coursebooks were used and one where the teacher-produced textbook was used will be reported.

#### **4.5 In-depth Illustration of a Class Under Observation Using Commercial Coursebook**

Observation of a class using 5 (commercial coursebook) for instance revealed how the theme of unit 3 “How can the eyes deceive the mind?” was “successfully taught” with the researcher-observer commenting that “set induction into the lesson was highly effective, the prompts in the book at the start of the unit were interesting, and classroom participation was excellent.” Students were very responsive to the teacher, who was enthusiastic, “eliciting responses from students and receiving loads of student responses.” The teacher had the support that was essential. There were so many prompts listed in the textbook, “excessive” in the words of the researcher observer, but despite that weakness, the teacher was seen as “selective, using only the ones that were best suited for the particular context” and “using some others of her own which were localized or relating more to Saudi students.” Lesson development was described by the researcher observer as “excellent.” All of these observation data confirmed and elaborated on the high ratings that these commercial coursebooks had in the nine items on the ELT-TEC instrument (1-9) under general features: (1. most of the tasks in the book are interesting; 2. tasks move from simple to complex; 3. task objectives are achievable; 4. cultural sensitivities have been considered; 5. the language in the textbook is natural and real; 6. the material is up to date; 7. it covers a variety of topics from different fields; and 8. the book contains fun elements (Table 2).

The teacher then moved on to lesson development, which was the main listening comprehension input. The input was from reputable sources such as National Geographic. Variety in the listening tasks was also evident as the other listening resource was from a history lecture. Because the video clips were

interesting, “especially the one from National Geographic,” the students were “absorbed in it while watching” and the “effects were clear in the discussion immediately following.” Such observations confirmed and elaborated on the high ratings that the commercial coursebooks obtained for the listening section, ELT-TEC, for item 10 “The book has appropriate listening tasks with well-defined goals,” item 12 “Tasks are efficiently graded according to complexity,” and item 13 “Tasks are authentic or close to real language situations.” One item, item 11 “Instructions are clear,” even scored very high (Table 2).

Much of the ease in language communication, even in the intermediate classroom like this one, was because “students seemed at ease with the vocabulary,” which supported the high rating of the commercial coursebooks in the vocabulary section, for item 23 “The load (number of new words in each lesson) is appropriate to the level,” and item 26 “Words are contextualized” (Table 2). The observations revealed that the book developers “were informed on the research in vocabulary,” so they put a lot of emphases on strategies for vocabulary load and distribution, which were important for acquisition and learning to take place— aspects such as “repetition and recycling of words.” What was even more encouraging was the reassurance that teachers had as each unit, “explicitly stating the vocabulary that was to be covered and extra information was provided to link the words to the *Oxford 3000* list.” Even parts of speech (v.) indicating verbs and (adv.) indicating adverbs were explicitly used, which the teachers thought were “extremely helpful” (TI). These data supported and elaborated on the high ratings from the ELT-TEC for item 24 “There is a good distribution (simple to complex) of vocabulary load across chapters and the whole book” (Table 2).

The classroom observation also revealed that words, such as deceive, deception, tricks, visual effects, infinite, camouflage, mimic, survival, and predator, were all known to the students before their reintroduction in this lesson (OFN). This supported the high rating on the ELT-TEC received for item 25 “Words are



efficiently repeated and recycled across the book” (Table 2). The repetition and recycling of words are important in learning, as new words introduced to the learner should be repeated at least seven times in intervals across the book (Thornbury, 2002). The only word in the listening input which was never covered prior to this lesson was ‘mirage,’ and the teacher “negotiated with learners through prompts the meaning of the word, using other words similar in meaning like ‘deception’ (which was already introduced in previous lessons) within the processes of negotiation” (OFN). This negotiation once completed was then reinforced with visuals (which were plentiful in the textbook and in the supplementary online material). This supported the high rating for item 26 “Words are contextualized on the ELT-TEC” (Table 2).

#### **4.6 In-depth Illustration of One Class under Observation Using the Teacher-Produced Textbook**

The teacher-produced textbook was used in an intermediate class. This unit had ‘obesity’ as the unit title. The teacher began the class with set induction routines. The teacher was chatty, and students were responsive. The teacher elicited responses such as “What leads to obesity?,” and the students responded well. Then the teacher moved into lesson development, and he got the students to work on pre-reading. The pre-reading had nothing to do with the concepts or ideas related to the passage that the students were about to read. It was “a boring multiple-choice exercise getting students to guess the meanings of words that would be introduced in the reading passage.” This observation supported and elaborated upon the moderate rating on the ELT-TEC for item 9 “The book has fun elements” (Table 2).

The next stage was the reading passage, which the teacher described as “heavy and boring” (TI), and the students described it as “uninteresting” (SI), as it was “all about the definition of obesity and its causes” (SI). This again supported the moderate rating for item 9 “The book has fun elements” (Table 2). Surprisingly,

some of the students said that there were “better things to read about than obesity on the Internet,” and they preferred to read about “how people overcame their condition” or “dealt with abuse from peers” as a result of their condition (SI).

The next stage of the lesson was when the reading was done and the students answered comprehension questions. The class concluded with the teacher checking the answers to the comprehension questions with the students. As there were still at least ten minutes of class time left, the teacher “kept checking his watch, looking lost.”

Another important observation recorded in this class, where teacher-produced textbook was used, was that the teacher seemed to “want to stretch the reading passages, wanting to exploit it beyond capacity by trying to lead students to do more discussion.” However, the “one-word or phrase-level responses the teacher got from the students” aborted any attempt at staying with the text. This confirmed the moderate rating obtained for all three items (item 14 “Activities are developed to initiate meaning communication,” item 15 “Activities are balanced between individual response, pair work, and group work,” and item 16 “Activities motivate students to talk” (Table 2). The absence of tasks that encouraged participation that involved pair work and group work led the teacher to doing all of the prompting. The teacher again confirmed the moderate rating received for item 16 “Activities motivate students to talk” (Table 2), by stating that “materials that do not stimulate learners’ imagination nor engage them cannot be exploited for collaborative work” (TI).

#### **4.7 Teachers’ Coping Strategies: Would This Have Influenced the Perceptions of Textbooks?**

This section reports the results related to the third research question:

RQ 3: How do teachers cope with shortcomings of materials in the classroom and how does this influence their perceptions of textbooks?

Unlike the predictive evaluations of materials, usually that which concerns the suitability of materials for selection purposes, the retrospective evaluation of materials takes into account factors such as the capacity of teachers in handling the challenges that come with materials use. This usually provides avenues for some aspects of the adaptation of materials, which may also influence the teachers' perceptions of books.

All 19 teachers believed that while the commercial coursebooks were superior in many aspects, there were challenges when it came to use of these books as well. These teachers also believed that “how they coped with the challenges and how successful they were in managing them also contributed to positive or negative evaluations at the macro level” (TI). Some accounts of how the teachers coped even in adverse teaching situations have been documented even in the Arab context elsewhere (Al-Subaiei, 2017).

The teachers stated that the “many similarities in themes” (TI) across the ten books meant that “advance level students, especially in the health track (from which some students went on to medical school), found them boring” (TI). The analysis of the textbooks carried out by the researchers did reveal that there was “intense repetition of words which was deployed as a strategy” (OFN), and this would have “limited the scope of the themes” so that this strategy (i.e., repetition) could be realized. The student interviews revealed that while those in the science and humanities track were “comfortable with the themes and found them interesting” (SI), those in “the highly competitive health and engineering track” (TI) found the themes “unable to motivate them and were not challenging enough” (SI). The interviews with the students also revealed that many of those in the health and engineering track “studied in international schools” or their parents “enrolled

them into extra English language tutorials” (SI) so that they could be fluent in the language. They were “competent users of the language” (OFN) and “needed more challenge” (TI). The teachers observed that teaching using the commercial coursebooks in the health and engineering track skipped “some parts of the units because the students found the reading too easy” (OFN). At the interviews it was confirmed that “omitting large chunks of materials in the advanced classes was possible because of the many options available with commercial coursebooks”—the teachers claimed the commercial coursebook learning systems (including the online supplementary materials) “were like a never-ending supply of materials” (TI). This teacher strategy of skipping or omitting unsuitable materials (whilst still having an abundance of materials for selection) confirmed the high rating that the commercial coursebooks scored on the ELT-TEC. Skipping or subtraction is considered an important part of the adaptation of teaching materials (Masuhara, 2022). The teachers teaching using the teacher-produced textbooks did not have the luxury of alternate resources as they lacked the huge array of supplementary materials that the commercial coursebooks came with; and since “omission and skipping ineffective materials were never an option” (TI) with the teachers using the teacher-produced textbooks, they struggled with the books, which could explain the moderate ratings, especially for listening (items 10-13) and speaking (items 14-16) (Table 2).

## **5. Conclusion**

This study reveals glaring weaknesses in the teacher-produced textbooks, especially in the areas of the outlay of supplementary materials, and listening and speaking. The major problems with the weaknesses in these areas led to, as the micro-evaluations revealed, the absence of motivation on the part of the students due to the lack of variety in materials use, which is considered by experts to reduce student engagement (Tomlinson, 2011). Originally these teacher-produced textbooks were meant to be supplementary books to gradually immerse preparatory year students to ESP or EAP. However, the expert validators were of

the opinion (like the majority of teachers interviewed) that the disciplines they covered were too wide (for ESP) and that the themes were too general. The experts were also of the opinion that all of the teacher-produced textbooks were like regular general English textbooks having regular reading comprehension exercises (not ones specific to ESP/EAP). They did not conform to the requirements for materials development for ESP (Hutchinson & Waters, 1987) or for EAP (Jordan, 1997).

There will be questions asked of decision-makers, implementers, and materials developers with regard to the relevance of having teacher-produced textbooks as components of a preparatory year English program. While they were originally meant to immerse students into ESP/EAP, there was no evidence that the transition succeeded, which in turn defeats the purpose of teaching ESP/EAP. This evaluation revealed that these teacher-produced textbooks were written in an ad hoc manner and in haste, and therefore their quality was affected. Many of the teachers in this study were even of the opinion that the teacher-produced textbooks must be removed from the system.

## **6. About the Authors**

Abdulaziz Alfahaid is a Professor of Applied Linguistics in the English language Department at Imam Abdulrahman Bin Faisal University (Saudi Arabia). He obtained his M.A. and Ph.D. in Applied Linguistics and TESOL from the School of Education at the University of Leicester, UK. He has taught applied linguistics and EFL and consulted on assessment, program evaluation projects, and pedagogy in ESL and adult education settings. His research focuses mainly on English language education, English for academic purposes, developing and evaluating preparatory year curriculum and programs, and student transition. He can be reached at [aalfahaid@iau.edu.sa](mailto:aalfahaid@iau.edu.sa).

Nada Alkhatib is currently an Associate Professor of Applied Linguistics in the English Language Department at Imam Abdulrahman Bin Faisal University in Saudi Arabia. She teaches academic writing, phonetics, phonology, and research methods in linguistics. She obtained her M.A. and Ph.D. in Applied Linguistics and TESOL from the University of Essex, UK. She can be reached at [nalkhatib@iau.edu.sa](mailto:nalkhatib@iau.edu.sa).

Jayakaran Mukundan is an expert in the area of materials development. His materials evaluation software has won Gold Medals at the British Invention Show, London (2009) and at IENA, Nuremberg, Germany (2010). He has also developed an online textbook evaluation checklist [www.textbookevaluation.com](http://www.textbookevaluation.com). He is currently a professor and senior research fellow at Taylor's University, Lakeside Campus, Malaysia.

## 7. References

- Ahour, T., & Ahmadi, E. (2012). Retrospective evaluation of textbook "Summit 2B" for its sustainability for EFL undergraduate students. *Journal of Educational and Social Research*, 2(5). 195–202.  
<https://www.richtmann.org/journal/index.php/jesr/article/download/11950/11555/43737>
- Al-Subaiei, M. S. (2017). Challenges in mixed ability classes and strategies utilized by ELI teachers to cope with them. *English Language Teaching*, 10(6), 182–189. <https://doi.org/10.5539/elt.v10n6p182>
- Barnard, R., & Randall, M. (1995). Evaluating course materials: A contrastive study in text book trialling. *System*, 23(3), 337–346.  
[https://doi.org/10.1016/0346-251X\(95\)00021-B](https://doi.org/10.1016/0346-251X(95)00021-B)
- Bixby, J., Brooks, M., Caplan, N., Craven, M., Daise, D., Douglas, S., Earles-Carlin, S., Freire, R., Gramer, M., Jones, R., McVeigh, J., Lynn, S., McClurem K., Norliff, C., Scanlon, J., Sherman, K., Vargo, M., Ward, C., Zwier, L., ...

- Hamilton, K. (Eds.). *Q: Skills for Success* (Level intro – 5). Oxford University Press.
- Boxer, D., & Pickering, I. (1995). Problems in the presentation of speech acts in ELT materials: The case of complaints. *ELT Journal*, 49(1), 44–58.  
<https://doi.org/10.1093/elt/49.1.44>
- Chao, T. C. A. (2010). Retrospective evaluation of a local college English textbook from an intercultural perspective. *Studies in English for Professional Communications and Applications*, 89, 57–61.
- Davari, H., Iranmehr, A., & Erfani, S. M. (2013). A critical evaluation of PNU ESP textbooks. *Journal of Language Teaching and Research*, 4(4), 813–823.  
<https://doi.org/10.4304/jltr.4.4.813-823>
- Ebadi, S. & Farjad, N. (2015). Evaluation of EAP textbooks: A comparison between SAMT English for medical students and Oxford English for careers: Medicine. *Journal of Applied Linguistics and Language Research*, 2(3), 133–149.  
[https://www.jallr.com/index.php/JALLR/article/view/48/pdf\\_46](https://www.jallr.com/index.php/JALLR/article/view/48/pdf_46)
- Ellis, R. (1997). The empirical evaluation of language teaching materials, *ELT Journal*, 51(1), 36–42. <https://doi.org/10.1093/elt/51.1.36>
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics: And Sex and Drugs and Rock “N” Roll* (4th ed.). Sage.
- Gholami, R., Noordin, N., & Rafik-Galea, S. (2017). A thorough scrutiny of ELT textbook evaluations: A review inquiry. *International Journal of Education and Literacy Studies*, 5(3), 82–91.  
<https://doi.org/10.7575/aiac.ijels.v.5n.3p.82>
- Gilmore, A. (2004). A comparison of textbooks and authentic interactions. *ELT Journal* 58(4), 363–374. <https://doi.org/10.1093/elt/58.4.363>
- Hartle, S. (2022). Developing blended learning materials. In. J. Norton & H. Buchanan (Eds.), *The Routledge handbook of materials development for language teaching* (pp. 399–413). Routledge.

- Hutchinson, T., & Waters, A. (1987). *English for specific purposes*. Cambridge University Press.
- Jordan, R. R. (1997). *English for academic purposes*. Cambridge University Press.
- Kashoob, M. M. D. (2018). *Evaluating ELT materials with specific reference to colleges of applied sciences general foundation program in Oman: Towards a viable checklist* (Unpublished doctoral dissertation). University of York, UK.
- Kirkgoz, Y. (2011). An evaluation of English textbooks in Turkish primary education: Students' and teachers' perceptions. *Eurasian Journal of Educational Research*, 44, 167–184. <https://ejer.com.tr/wp-content/uploads/2021/01/ejer.2011.44.10.pdf>
- Lisna, M. (2016). Retrospective evaluation of ELT coursebook effectiveness. *Advanced Education*, 19, 87–92. <https://doi.org/10.20535/2410-8286.81589>.
- Littlejohn, A. (1998). The analysis of language teaching materials: Inside the trojan horse. In B. Tomlinson (Ed.), *Materials development in language teaching* (pp. 190–216). Cambridge University Press.
- Littlejohn, A. (2022). The analysis and evaluation of language teaching materials. In J. Norton & H. Buchanan (Eds.), *The Routledge Handbook of Materials Development for Language Teaching* (pp. 179–211). Routledge.
- Maryam, D.T. & Sara, H. (2014). Evaluation of ESP textbooks: Evidence from ESP Textbooks of Computer Engineering Majors. *International of Research Studies in Language Learning*, 3(2), 55–68. <https://doi.org/10.5861/ijrsll.2013.469>
- Masuhara, H. (2022). Approaches to materials adaptation. In J. Norton & H. Buchanan (Eds.), *The Routledge handbook of materials development for language teaching* (pp. 277–290). Routledge.
- McDonough, J., & Shaw, C. (2003). *Materials and methods in ELT* (2nd ed.). Blackwell Publishing.



- McGrath, I. (2002). *Materials evaluation and design for language teaching*. Edinburgh University Press.
- 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* (pp. 336–352). Continuum International Publishing Group.
- Mukundan, J. (2007). Evaluation of English language textbooks: Some important issues for consideration. *Journal of NELTA (Nepal English Language Teachers' Association)*, 12(1&2), 80–84.  
<https://www.nelta.org.np/uploads/upload/M1CnoN.pdf>
- Mukundan, J. (2010). *ESL textbook evaluation: A composite framework*. LAP Lambert Academic Publishing.
- Mukundan, J., & Aziz, A. (2009). Loading and distribution of the 2000 high frequency words in Malaysian English language textbooks for form 1 to form 5. *Pertanika Journal of Social Science and Humanities*, 17(92), 209–220.
- Mukundan, J., & Kalajahi, S. A. R. (2013). Evaluation of Malaysian English language teaching textbooks. *International Journal of Education and Literacy Studies*, 1(1), 38–46.  
<http://dx.doi.org/10.7575/aiac.ijels.v.1n.1p.38>
- Mukundan, J., & Nimehchisalem, V. (2012). Evaluative criteria of an English language textbook evaluation checklist. *Journal of Language Teaching Research*, 3(6), 1128–1134. <https://doi.org/10.4304/jltr.3.6.1128-1134>
- Mukundan, J., & Roslim, N. (2009). Textbook representation of prepositions. *English Language Teaching*, 2(4), 13–24.
- Nemati, A. (2009). Evaluation of an ESL English coursebook: A step towards systematic vocabulary evaluation. *Journal of Social Sciences* 20(2), 91–99.  
<https://doi.org/10.31901/24566756.2009/20.02.02>

- Nimehchisalem, V., & Mukundan, J. (2015). Refinement of the English Language teaching textbook evaluation checklist. *Pertanika Journal of Social Sciences & Humanities*, 23(4), 761–780.
- Prabhu, N. S. (2019). Foreword: on the value of shared experiences. In A. Maley (Ed.), *Developing expertise through experience* (pp. 3–7). British Council.
- Richards, J. C., & Lockhart, C. (1994). *Reflective Teaching in Second Language Classrooms*. Cambridge University Press.  
<https://doi.org/10.1017/CBO9780511667169>
- Richards, J. C., & Rodgers, T. (1982). Method: Approach, Design, and Procedure. *TESOL Quarterly*, 16(2), 153–168. <https://doi.org/10.2307/3586789>
- Skierso, A. (1991). Textbook selection and evaluation. In M. Celce-Murcia (Ed.), *Teaching English as a Second or Foreign Language* (pp. 432–453). Heinle & Heinle.
- Tang, X., & Zheng, D. (2018). A retrospective evaluation of college ELT textbooks in China. *Proceeding of the 2018 International Conference on Social Science and Education Reform (ICSSER 2018)*, 57–61.  
<https://doi.org/10.2991/icsser-18.2018.13>
- Thornbury, S. (2002). *How to Teach Vocabulary*. Pearson Education Limited.
- Tomlinson, B. & Masuhara, H. (2021). *SLA applied: Connecting theory and practice*. Cambridge University Press.
- Tomlinson, B. (2011). *Material development in language teaching* (2nd ed.). Cambridge University Press.
- Ton Nu, A. T., & Murray, J. (2020). Pragmatic content in EFL textbooks: An investigation into Vietnamese national teaching materials. *The Electronic Journal for English as a Second Language*, 24(3), 1–28.
- Usó-Juan, E. (2008). The presentation and practice of the communicative act of requesting in textbooks: Focusing on modifiers. In: Soler, E. A., Jordà, M. S. (Eds.) *Intercultural Language Use and Language Learning*. Springer.  
[https://doi.org/10.1007/978-1-4020-5639-0\\_12](https://doi.org/10.1007/978-1-4020-5639-0_12)

Zarifi, A., & Mukundan, J. (2012). Phrasal verbs in Malaysian ESL textbooks, *English Language Teaching* 5(5), 1–9.  
<http://dx.doi.org/10.5539/elt.v5n5p9>

Zarifi, A., & Mukundan, J. (2015). A corpus-based study of semantic treatment of phrasal verbs in Malaysian ESL secondary school textbooks. *Pertanika Journal of Social Science and Humanities*, 23(4), 793–808.