Systematic Literature Review on Evaluation of English Language Textbooks: A Decade of Research

Yaxiong Li[1,2], Ramiza Haji Darmi[1], Yap Ngee Thai[1]

[1] Department of English, Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Selangor, Malaysia

[2] Department of Student Affairs, Hunan Institute of Technology, Hengyang, Hunan, 421002, China

Correspondence: Ramiza Haji Darmi, Department of English, Faculty of Modern Languages and Communication, Universiti Putra Malaysia, 43400, Serdang, Selangor, Malaysia. E-mail: ramiza@upm.edu.my

Abstract

Textbooks, essential for teaching and learning, have become a burgeoning research subject in education. However, textbook evaluation has not garnered adequate attention in the context of English Language Teaching (ELT). This hinders the identification of key characteristics, focused learning themes, and gaps encountered within the EFL/ESL educational landscape. This systematic literature review, employing the ROSES framework, explores key characteristics and learning themes within English language textbook evaluation in the ELT context, identifying suggestions for future evaluations. The review involves searching, screening, evaluating, and synthesizing pertinent articles published in the last decade, from 2014 to 2023, across Scopus and Web of Science databases. Out of 2304 articles identified through a search of keywords including "textbook" and "evaluation" alongside their synonyms, 30 studies meeting the inclusion criteria are shortlisted after the quality appraisal using MMAT. The review finds that questionnaires, surveys and checklists were the most common methods used for ELT textbooks. Based on the findings from the review, this paper discusses a wide range of indicators or criteria involved in evaluating ELT textbooks, particularly in the evaluation of culture and pragmatics. Our research has revealed a need for in-depth exploration using qualitative and mixed-methods approaches, emphasizing the necessity for broader comparative studies and a more diverse range of perspectives in educational assessments to bridge knowledge gaps. This study suggests that further research on textbook evaluation in the ELT context is still necessary.

Keywords: textbook, textbook evaluation, English language, English as a foreign language, systematic literature review

1. Introduction

Textbooks are essential tools for curriculum delivery, organizing content, and enhancing the teaching and learning process (Benabed & Mehdaoui, 2023; Shi, 2022; Syairofi et al., 2022). Textbooks, considered the cornerstone of English as a Foreign Language courses (Sheldon, 1988), serve a dual role. They provide learners with authentic linguistic materials and essential information (Ahour et al., 2014; Goodarzi et al., 2021) while also acting as crucial guides for novice teachers at the beginning of their careers.

Due to the substantial role of English textbooks, evaluating textbooks is a crucial component in English learning and teaching (ELT) context (Goodarzi et al., 2021; Maleki et al., 2014; Rangaraju, 2020). Textbook evaluation ensures the selection and use of appropriate textbooks (Hamidi et al., 2016) that meet learners' needs and align with the program's objectives and methods towards facilitating effective teaching and learning (Cunningsworth, 1995). A precise understanding of textbook evaluation in the ELT context is lacking (Benabed & Mehdaoui, 2023; Shi, 2022; Simsek, 2022), including a descriptive analysis of related empirical research and a comprehensive framework of how to conduct textbook evaluation in the ELT context.

This study aims to investigate the key features and applied research methods in the ELT context. The objective is to identify evaluation aspects, research gaps and potential methodologies that could be incorporated into future textbook evaluations in teaching English as a Foreign Language (EFL) or English as a Second Language (ESL) areas. The rationale, therefore, prompted the main research question: "How do ELT textbooks users evaluate the textbooks?". Subsequently, the following research questions were identified:

1) What are the primary research methods employed in the selected studies?

2) What are the main objectives of the selected studies?

3) What are the outcomes of the studies?

4) What are the research gaps?

5) What are the suggestions for future ELT textbook evaluation?

2. Method

This section details the systematic review methodology, with initial searches conducted from April 26 to May 5, 2024, followed by these

steps:

2.1 Review Protocol---ROSES

The systematic literature review (SLR) adopted Reporting Standards for Systematic Evidence Syntheses (ROSES) framework developed by Haddaway et al., (2018). The main objective of ROSES is to promote transparency and ensure quality control throughout the review process, acting as a stringent scientific approach to gain insights within a specific research domain. The review initiated with the formulation of research questions based on the PICo method. Subsequently, the study went through a systematic search strategy comprising three phases: identification, screening, and eligibility. Next, quality appraisal based on criteria adapted from Hong et al. (2018) was applied to assess the selected articles before their inclusion in the review. Finally, selected articles went through data extraction and analysis. Data extraction followed the primary research question, while thematic synthesis for quantitative and qualitative data guided the analysis process.

2.2 Formulation of Research Questions

The authors applied the PICo mnemonic, representing "P" (Population or Problem), "I" (Interest), and "Co" (Context) to formulate research questions (Lockwood et al., 2015). Then, the three primary facets were integrated into the review, including the users (Population), evaluation of the teachers' and students' textbook (Interest), and EFL/ESL (Context). This approach aids in formulating the central research question for the study: "How do ELT textbooks users evaluate ELT textbooks?".

2.3 Systematic Review Process

Three systematic procedures suggested by Shafril et al. (2018) were utilized to acquire relevant articles. These processes include identification, screening, and eligibility. The processes enabled the researchers to systematically locate and integrate studies, facilitating the execution of a well-organized and transparent SLR.

2.3.1 Identification

The first process is identifying keywords from the defined research inquiries. The three primary keywords are textbook, evaluation, and users. To enrich this terminology, the authors explored synonyms, related expressions, and variations through online resources like thesaurus and online dictionary. Also, the terms included keywords from past studies, suggestions from Scopus, and insights from experts in this domain. This iterative process led to the identification of keywords synonymous with "book", "evaluate", "assess", "user", "teacher", and "student". These keywords were combined and processed using various search functions, including field code functions, phrase searching, wildcards, truncation, and Boolean operators, across two databases: Scopus and Web of Science (refer to Table 1). These databases were chosen for their significant impact in the field of social science and accessibility to institutions and scientific associations. In total, 12,187 potential articles were identified from these databases.

Table 1. Search String used in the Selected Databases

Database	Search String	Hits	Date
Scopus	TITLE-ABS-KEY ((textbook* OR book*) AND (evaluat* OR assess*) AND (teacher* OR student* OR user*) AND (e*l))	10566	26/4/2024
Web of Science	textbook* OR book* (Topic) and evaluat* OR assess* (Topic) and teacher* OR student* OR user* (Topic) and E*L (Topic)	3218	26/4/2024

2.3.2 Screening

The second process is screening that involves the inclusion and exclusion of articles. It is a process executed with the aid of the database conducted by the first author. The selection process adhered to a specific set of criteria outlined in Table 2. In line with the concept of "research field maturity", highlighted by Kraus et al. (2020), this review focuses on articles published between 2014 and 2023. This time frame was chosen to ensure an adequate number of published studies for a comprehensive and representative review, aligning with the latest trends of textbook evaluation within ELT context. It is worth noting that, for clarity, only articles on textbooks used for EFL or ESL and teaching were considered. Also, to ensure the high quality of articles to be analyzed, review articles, chapters of a book, a book, and conference proceedings are not included. Considering the SLR's emphasis on textbook evaluation in teaching ESL/EFL, research studies in social science were preferred.

Inclusion criteria and exclusion criteria were applied to ensure that the selected studies could effectively address the research questions outlined above. The inclusion criteria are organized into four categories, as outlined in Table 2.

Table 2. Inclusion and Exclusion Criteria

Criterion	Inclusion	Exclusion	Excluded articles
Textbook type	English textbook	Non-English textbook	
Timeline	2014-2023	2013 and earlier	4042 (Scopus)+842 (WOS)
Document type	Journal Articles	Review articles, chapters in a book, a book, conference proceedings, etc.	4341 (Scopus)+731 (WOS)
Language	English	Non-English	280(Scopus)+166(WOS)
Subject area	Context of ELT in Social Science	Other non-social science studies	779(Scopus)

A total of 2304 papers were retrieved after entering the title screening phase. All raw results were compiled in an Excel repository. At this stage, bibliometric measurements, such as the number of citations or journal rankings, were not taken into consideration. Firstly, a spreadsheet was maintained in EXCEL format. Then, the resulting studies were added to another sheet of the spreadsheet and analyzed based on the title in accordance with the inclusion criteria (ICs) or exclusion criteria (ECs) during the screening phase. 132 duplicates were removed before the title screening. Then, in the title screening phase, 2172 articles were removed from the spreadsheet.

2.3.3 Eligibility

The process is done manually by reading the titles, keywords, and abstracts of the selected papers. This is to determine their relevance to textbook evaluation and ELT context. The selection process focused solely on English language textbooks, resulting in a narrowed set of 61 articles chosen for in-depth review following an initial screening of all available abstracts. Of these, 13 articles were removed because it focused on the criteria for textbook evaluation instead of evaluating certain ELT textbook(s) and one article was removed for its inaccessibility. Fourthly, the author promptly evaluated the entire content of the articles to determine their eligibility for full-text assessment. The number of papers selected after reading the full text is 57. During this phase, an expert in ELT area was invited to review the search process and the eligibility of the selected articles. Finally, a total of 44 articles were selected at the quality appraisal stage.

2.3.4 Quality Appraisal

The Mixed-Method Appraisal Tool (MMAT) developed by Hong et al. (2018) served as the framework for assessing the methodology and analysis of the selected studies during the quality appraisal phase. MMAT is designed to evaluate systematic mixed studies reviews across various research designs, including qualitative research, randomized controlled trials, non-randomized studies, quantitative descriptive studies, and mixed methods studies. With guidance from the corresponding author, each article underwent meticulous scrutiny, focusing on its methodology and analysis. Following two screening questions applicable to all research types, the articles were evaluated against five criteria specific to their research design, with responses categorized as "Yes," "No," or "Can't tell." Articles meeting at least three criteria were considered for inclusion in the review. Decisions were made through consensus between the authors. Following this process, both authors agreed that the selected articles met the minimum quality standards for methodology and analysis. Out of the total, 30 articles met all criteria. The complete details are presented in Appendix A. 14 articles were excluded after the quality appraisal and the results are presented in Table 3. Additionally, Figure 1 shows the flow diagram that illustrates the search process.

Table 3. Results of the Quality Assessment

Authors/Year	S1	S2	Research design	QA1	QA2	QA3	QA4	QA5	Number of criteria	Inclusion or not
Benabed & Mehdaoui, 2023	✓	✓	MX	✓	✓	✓	✓	✓	5	✓
Syairofi et al., 2022	×	×	QL							×
Banaruee et al., 2023	✓	✓	QL	✓	✓	✓	✓	✓	5	✓
Shi, 2022	✓	✓	QL	✓	✓	✓	✓	✓	5	✓
Osokina, 2022	×	×	QL							×
Cao et al., 2022	✓	✓	QL	✓	✓	✓	✓	✓	5	✓
Goodarzi et al., 2021	✓	✓	MX	✓	✓	✓	✓	✓	5	✓
Simsek, 2022	✓	✓	MX	✓	✓	✓	✓	✓	5	✓
Nazim, 2021	×	×	MX							×
Soto & Cevallos, 2021	×	×	QL							×
Santos, 2020	×	×	QN(DC)							×
Vahdat et al., 2020	✓	✓	QN(DC)	✓	✓	✓	✓	✓	5	✓
Derakhshan, 2021	✓	✓	MX	✓	✓	✓	✓	✓	5	✓
Nu & Murray, 2020	✓	✓	MX	✓	✓	✓	✓	✓	5	✓
Cárcamo Morales, 2020	×	×	QL							×
Rangaraju, 2020	×	×	QL							×
Pérez-Hernández, 2019	×	×	QL							×
Dennis, 2018	✓	✓	QN(DC)	✓	✓	✓	✓	✓	5	✓
Zare-Ee & Hejazi, 2018	✓	✓	QN(DC)	✓	✓	✓	✓	✓	5	✓
Bouzid, 2017	✓	✓	MX	✓	✓	✓	✓	✓	5	✓
Safa et al., 2015	✓	✓	QN(DC)	✓	✓	✓	✓	✓	5	✓
Hamidi et al., 2016	×	×	QL							×
Aghazadeh, 2015	✓	✓	MX	✓	✓	✓	✓	✓	5	✓
Ahour et al., 2014	✓	✓	QN(DC)	✓	✓	✓	✓	✓	5	✓
Asiyaban & Zamanian, 2014	✓	✓	QN(DC)	✓	✓	✓	✓	✓	5	✓
Maleki et al., 2014	✓	✓	QN(DC)	✓	✓	✓	✓	✓	5	✓
Azarfam & Noordin, 2014	×	×	QL							×
Rashidi & Kehtarfard, 2014	✓	✓	QN(DC)	✓	✓	✓	✓	✓	5	✓
Namaghi et al., 2014	×	×	QL							×

Study	S	S	Type						QA	
Hessamy & Mohebi, 2014	✓	✓	MX	✓	✓	✓	✓	✓	5	✓
Haghighi, 2014	✓	✓	QN(DC)	✓	✓	✓	✓	✓	5	✓
Hooman, 2014	×	×	QN(DC)							×
Orfan et al., 2021	✓	✓	QN(DC)	✓	✓	✓	✓	✓	5	✓
Amiri & Rezvani, 2021	✓	✓	MX	✓	✓	✓	✓	✓	5	✓
Amerian & Tajabadi, 2020	✓	✓	MX	✓	✓	✓	✓	✓	5	✓
Bababayli & Kiziltan, 2020	✓	✓	QN(DC)	✓	✓	✓	✓	✓	5	✓
Karamifar et al., 2014	×	×								×
Bui & Nguyen, 2023	✓	✓	QL	✓	✓	✓	✓	✓	5	✓
Alharbi, 2022	✓	✓	QL	✓	✓	✓	✓	✓	5	✓
Li & Deocampo, 2021	✓	✓	MX	✓	✓	✓	✓	✓	5	✓
Vasheghani, 2020	✓	✓	QL	✓	✓	✓	✓		5	✓
Biria & Mehrabi Boshrabadi, 2015	✓	✓	MX	✓	✓	✓	✓	✓	5	✓
Ahmed et al., 2023	✓	✓	MX	✓	✓	✓	✓	✓	5	✓
Arshad et al., 2020	✓	✓	QL	✓	✓	✓	✓	✓	5	×
Mozejko, 2016	×	×								×

Note: S=screening questions (for all types). Further appraisal may not be feasible or appropriate when the answer is "No" or "Can't tell" to one or both screening questions. QA=quality assessment; QN(DC)=qualitative descriptive; QN(NR)=quantitative non-randomized; QL=qualitative; MX=mixed-method; C=can't tell
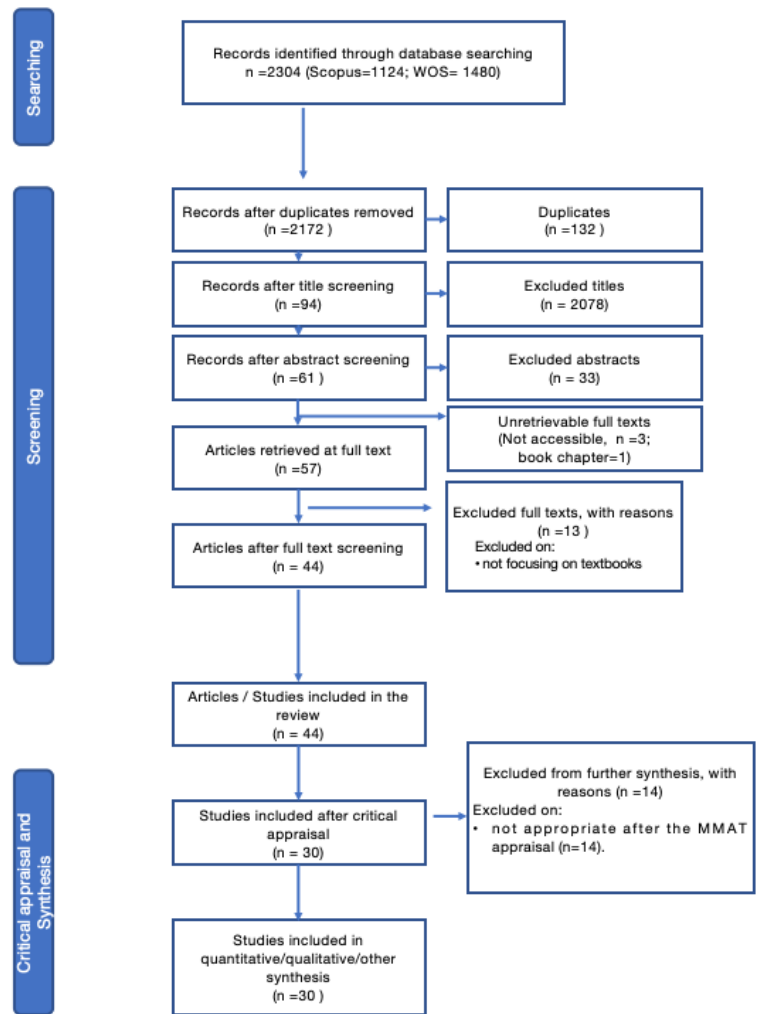

Figure 1. ROSES Flow Diagram for Systematic Review]

3. Results

30 relevant studies remained after the rigorous screening criteria. These articles went through the process of collection, assessment, and analysis. Finally, the SLR provides evidence on research methods, objectives, and further suggestions for textbook evaluation in the ELT context. It is important to note that specific intervention issues may have been addressed in other studies that were not covered in this review.

3.1 Quantitative Findings

The primary features of the research methods were examined, including the year of publication, level of education, countries involved, participant demographics, sample size, and types of methodologies.

3.1.1 Publication Year

Figure 2 provides an overview of the distribution of studies on textbook evaluations from 2014 to 2023. Year 2020 and 2014 reported the highest number of publications (n=6), followed by Year 2021 (n=5). Appendix B provides the complete list of the studies. Generally, the trend of research on textbook evaluation in ELT has fluctuated over the past decade.
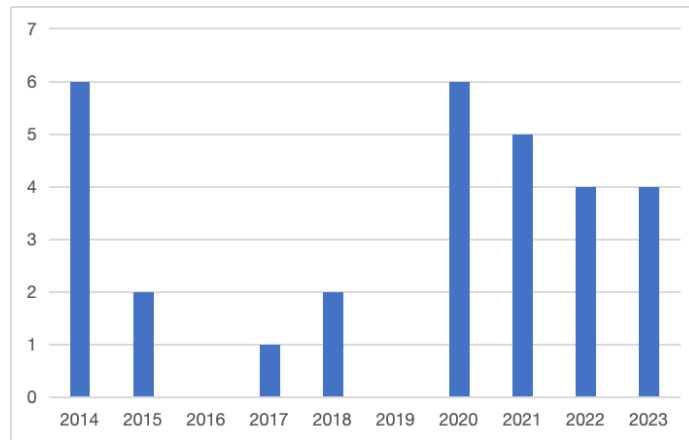

Figure 2. Year of Publication

3.1.2 Countries

Figure 3 illustrates the 12 countries and regions involved in the selected research on textbook evaluation, attracting interest from researchers globally. Iran leads in the number of publications (n=15), followed by Vietnam (n=3). Appendix C provides for the complete list of the studies.
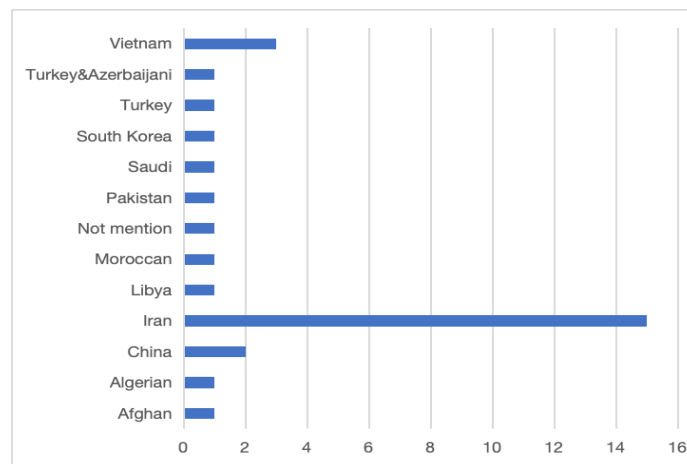

Figure 3. Country of Publication

3.1.3 Level of Education

Most studies (n=15) focused on middle and secondary education, followed by 11 studies targeting higher education. Solely one study addressed primary education, while three studies examined various educational levels. Appendix D provides the complete list of the studies. In respect to textbooks for middle and secondary education, only three studies specified participant ages ranging from 13 to 18 (Aghazadeh, 2015; Goodarzi et al., 2021; Rashidi & Kehtarfard, 2014). Among the evaluation of higher education textbooks, four studies (Haghighi, 2014; Dennis, 2018; Li & Deocampo, 2021; Safa et al. 2015; Simsek, 2022) reported learners aged between 18 and 30. Appendix D outlines the complete list of the different levels of education.

3.1.4 Subject

Figure 4 shows that 63.3% of the analyzed research utilized a single group of subjects comprising students (n=5), teachers (n=5), or researchers (n=9). Furthermore, 7 studies explored the perspectives of both students and teachers. Three studies examined the viewpoints of teachers-experts and teachers-students-experts regarding the content and construct validity of teaching materials in ELT (see Appendix E). Across these studies, criteria for selecting subjects included teaching experience and educational background. Two studies specified evaluators with at least more than 1 year of teaching experience (Ahour et al., 2014; Banaruee et al., 2023). Additionally, two studies indicated subjects having degrees from MA to PhD (Hessamy & Mohebi, 2014; Orfan et al., 2021). One study chose subjects with predetermined background information, such as native speakers (Safa et al., 2015).
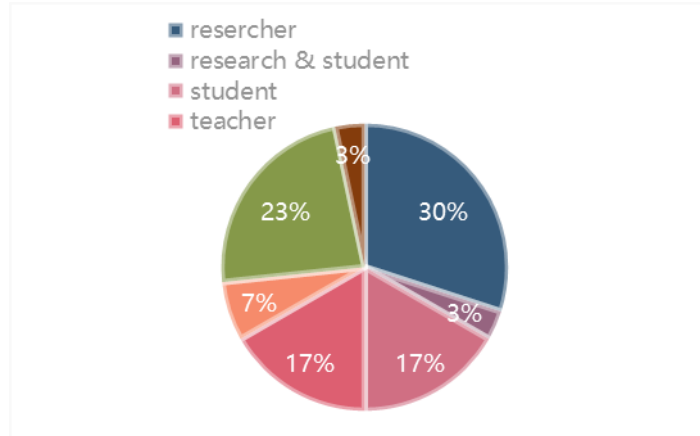

Figure 4. Subject of Publication

3.1.5 Sample Size of Subjects

Among the 30 studies, 11 studies (36.7%) had a sample size of over 100 participants, followed by 10 studies (33.3%) with a medium sample size of 11-100 participants. Furthermore, 2 studies (6.7%) had a sample size below 11, and 7 studies (23.3%) did not indicate the sample size of subjects (Appendix F provides the list of studies).

3.1.6 Research Methods

The mixed methods were the most frequently employed (n=12, 40%) approach, followed by the quantitative method (n=11, 35.5%) and the qualitative approach (n=7, 23.3%) (refer Appendix G). The primary data collection instruments include surveys, questionnaires and checklists. Out of the 30 studies, 13 (n=43.3%) utilized questionnaires, the most prevalent data collection instrument, followed by surveys (n=8, 26.7%) and checklists (n=8, 26.7%). The majority (n=17, 56.7%) of the studies used an adapted checklist to do the survey or develop questionnaires. (Appendix G provides the list of studies).

3.1.7 Aspects of Evaluation

More than half of the studies (n=17, 56.7%) evaluated the full ELT textbooks in general, while 13 studies (43.3%) evaluated pre-determined aspects of the textbook or the sub-genre or sub-theme of the textbook, including topics, activities, speech acts, and cultures.

Additionally, VOS Viewer was extracted to construct a map of links based on the co-occurrence of keywords in the abstracts and titles of each research article. Figure 5 illustrates the correlation among the keywords across the selected papers.

Based on the VOS viewer map, insights into the co-occurrence of keywords across 30 selected articles on ELT emerged. Firstly, an emphasis is placed on EFL context. Notably, expanding from this central cluster, connections to other research areas emerged, such as "culture representation" and "pragmatics", suggesting their relevance in evaluating textbooks for EFL learners. Secondly, the significance of needs analysis and curriculum design is underscored. Terms such as "content analysis" and "checklist" are intricately linked with evaluation, suggesting employing content analysis and checklists as prevalent methods for textbook assessment. Moreover, the frequent co-occurrence of "textbook", "English materials", "textbook evaluation", and "EFL" underscores the pivotal role textbooks play in EFL instruction and the imperative of their evaluation. Additionally, terms such as "intercultural competence," "culture," "disagreement", and "collocation" spotlight the cultivation of specific language proficiencies and knowledge pertinent to EFL learners, indicating a nuanced focus on language aspects. Similarly, "learners" and "teachers" are linked by cultural senses, indicating a focus on textbooks designed for users of ESL or EFL In summary, this examination of the selected articles encompasses diverse factors and elements relevant to textbook evaluation in EFL contexts, with an inclination towards pragmatic teaching methodologies over theoretical frameworks.

Figure 5. Overlay Visualization of the Selected Paper

3.2 Qualitative findings

3.2.1 Evaluation of English Textbooks from a Macro Perspective

The macro perspective aims to evaluate English textbooks by identifying their strengths and weaknesses across different regions, assessing whether the textbook meets the needs and suitability of its users or the program, and exploring the degree of alignment between the textbook and other theories.

After examining experts' opinions on content, grammar, vocabulary, attractiveness, activities and exercises, Benabed & Mehdaoui (2023) evaluated the suitability of the coursebook for third-year primary school learners based on their age, interest, and attractiveness. A criterion-based checklist, questionnaire, and survey are among the most popular assessment tools (McGrath, 2016; Richars, 2016; Shak, 2022). This is also proved in the 30 selected studies (Ahour et al., 2014; Amiri & Rezvani, 2021; Azarfam & Noordin, 2014; Benabed & Mehdaoui, 2023; Goodarzi et al., 2021; Hamidi et al., 2016, p. 3; Hooman, 2014; Li & Deocampo, 2021; Maleki et al., 2014; Manoochehr, 2013; Santos, 2020; Soto & Cevallos, 2021; Zare-Ee & Hejazi, 2018) in the past decade. Besides, a metaphor analysis (Shi, 2022) was carried out to evaluate an English for Academic Purposes (EAP) textbook by examining the beliefs of EFL graduate students in a Chinese university setting. The participants were required to write a paper in which they completed the metaphor "the academic English textbook is... because..." to express their attitudes toward the textbook. In addition to the aforementioned aspects, various considerations were taken into account, such as activities (Amerian & Tajabadi, 2020; Arshad et al., 2020; Banaruee et al., 2023; Derakhshan, 2021; Safa et al., 2015; Vahdat et al., 2020; Zare-Ee & Hejazi, 2018), subjects (Ahour et al., 2014; Bouzid, 2017), four language skills (Asiyaban & Zamanian, 2014;), authenticity (Maleki et al., 2014), illustrations (Haghighi, 2014), and culture (Aghazadeh, 2015; Orfan et al., 2021).

Additionally, conducting a needs analysis is crucial in textbook evaluation. Rashidi & Kehtarfard (2014) used questionnaires and distributed them to 180 third-grade female high school students in Iran to assess their perceived foreign language needs, serving as the foundation for textbook evaluation. The findings revealed that while the majority of students deemed all language skills and components important, the textbook did not adequately support all of them. Amiri & Rezvani (2021) investigated the vertical and horizontal alignment of educational objectives. Drawing upon Bloom's revised taxonomy, the study found that textbooks accommodated lower-order knowledge types and cognitive skills instead of higher-order levels. Simsek (2022) explored the flexibility of English for Specific Purposes (ESP) materials through the Inputs-Processes-Outcomes (IPO) model, focusing on corpus and consumer insights from the Turkish EFL context. It also suggested that ESP materials in the dental English context need transformation to achieve deeper learning outcomes, such as learner-compiled portfolios, increased visibility for women dentists, creative use of illustrations, attention to grammar, and integration of cross-cultural elements.

3.2.2 Evaluation of ELT Textbooks from Subthemes

Researchers have investigated different aspects or subthemes of the textbook, among which the culture aspect (Amerian & Tajabadi, 2020; Arshad et al., 2020; Banaruee et al., 2023; Derakhshan, 2021; Safa et al., 2015; Vahdat et al., 2020) turns out to be the most popular one, especially in the recent five years. Culture, crucial for creating connections between newly received information and prior knowledge (Banaruee et al., 2023), may have misconceptions of the target culture and inadequacy of the local culture (Vahdat et al., 2020). Teachers, students and educators need to be more aware of the inter-cultural knowledge, practices, and skills (Safa et al., 2015).

Besides, collocations (Cao et al., 2022), readability (Dennis, 2018) and pragmatic components (Bababayli & Kiziltan, 2020; Bui & Nguyen, 2023; Nu & Murray, 2020) are also investigated in the 30 selected studies.

4. Discussion

4.1 Suggestions and Recommendations for Future Research

For future studies in the field of textbook evaluation within ELT context, scholars have offered insightful theoretical and pedagogical suggestions. At the theoretical level, significant aspects of future studies include how to evaluate teachers' actual application of textbooks for student achievement and classroom discourse (Simsek, 2022); how various methods of data collection are needed such as interview (Bui & Nguyen, 2023; Derakhshan, 2021; Orfan et al., 2021; Shi, 2022); and how different evaluators will influence the textbook evaluation (Ahour et al., 2014; Bui & Nguyen, 2023; Hamidi et al., 2016; Hessamy & Mohebi, 2014; Hooman, 2014). Another suggestion is to use a different checklist or questionnaire and re-test its effectiveness and usefulness for textbook evaluation (Vahdat et al., 2020).

Furthermore, research has suggested new supplementary materials to develop and revise the textbook. Pedagogically, since there is no ideal textbook, almost all researchers found that there are some deficiencies that need to be identified and improved (Benabed & Mehdaoui, 2023). These deficiencies include well-defined headings and subheadings, content (Haghighi, 2014; Asiyaban & Zamanian, 2014; Benabed & Mehdaoui, 2023; Maleki et al., 2014; Nu & Murray, 2020; Rashidi & Kehtarfard, 2014), such as cultural knowledge (Arshad et al., 2020; Banaruee et al., 2023; Safa et al., 2015), and collocations (Cao et al., 2022), authentic materials (Ahmed et al., 2023; Amerian & Tajabadi, 2020; Bababayli & Kiziltan, 2020; Bouzid, 2017; Goodarzi et al., 2021), simple grammar and activities (Amiri & Rezvani, 2021); and topics (Goodarzi et al., 2021).

These collective recommendations highlight the dynamic nature of textbook evaluation in ELT, urging scholars and educators to adopt a multifaceted approach integrating theoretical insights, practical strategies, and methodological innovation.

4.2 Gaps in the Past Studies on Textbook Evaluation in ELT Context

This review has identified several gaps in research regarding textbook evaluation in teaching ESL/EFL. These gaps call for future studies in this field.

Checklists, questionnaires, and surveys have been the main methods used in textbook evaluation in the ELT context to collect qualitative and quantitative data. Therefore, future ELT research would be more convincing if the researchers exploit different research methodologies or combine the research instruments, for instance questionnaire and interview. Additionally, the evaluation of textbooks across varying educational levels and regional contexts presents a commendable initiative. Shi's (2022) suggestion to incorporate the perspectives of multiple evaluators would enrich the evaluation process. Thus, rich findings are expected from every study. The past studies focused on two aspects: evaluation of printed textbooks from a macro perspective and from the sub-themes. They have not treated technology-based materials, including multi-modal materials, e-textbooks and Artificial Intelligence (AI), in line with the development of technology. The technology-based materials evaluation and the socio-linguistic aspects of materials are the leading research directions related to textbook evaluation in ELT context.

Textbook evaluation is essential not only at the theoretical level but also at the practical and pedagogical level, which needs more experimental and non-experimental studies. Future works must conduct research based on the opinions of education officers, teachers, students, and experts in language learning and teaching concerning a wider theoretical framework, multifaceted approach, and methodology of textbook evaluation in the ELT context.

4.3 Generalizability of Findings to Other Contexts

As presented earlier, the majority of studies were conducted in Iran. Thus, concerns may arise regarding the generalizability of textbook evaluations to other cultural settings and contexts. To mitigate this, future research could assess textbooks across various subjects and countries, taking into consideration factors like internet connectivity, educational levels, and users' demographics. This is in accordance with Vasheghani (2020)'s research. Analyzing similarities and differences in educational contexts could provide insights into how these variations impact users' textbook evaluations. Future studies may possibly delve into the substantial roles of culture and second language acquisition (SLA) played in textbook evaluation.

5. Conclusion

Research on the evaluation of ELT materials is continually advancing, and there is an increase in the number of studies using diverse research methods. Despite the growing number of studies and the expanding scope of research methods encompassing various aspects of teaching materials, numerous challenges persist. This paper uses the SLR technique to provide an overview of the existing research efforts and obstacles related to textbook evaluation in ELT context. Similarities exist between the methodologies analyzed in the selected studies. This study obtains three main findings.

Firstly, the number of papers dedicated to textbook evaluation in ELT context is limited. As shown above, in the last decade, a total of 2304 articles were published in Scopus and Web of Science, with only 46 being research papers on ELT materials, and 30 meeting the strict criteria, constituting a mere 1.3% percent. Moreover, regional imbalances in publications exist, with Iran leading in the number of publications among all countries, contributing 15 out of the 30 selected papers, amounting to 48.3%. This might be attributed to its poor network connectivity and heavy reliance on printed textbooks, as the primary source of firsthand English language learning material. Additionally, despite the increasing number of studies on textbook evaluation in China and other regions, the studies represent a relatively small proportion of the total number of papers focusing on foreign language, failing to align with the significance of teaching materials in

education.

Secondly, a single, one-sided research perspective has been applied. Relevant research mainly adopts two perspectives: one involves evaluating the in-use textbooks from users' perspectives through a checklist or questionnaire, and the other is to discuss the sub-themes of the textbook. More studies tend to concentrate on specific language skills (Banaruee et al., 2023; Zare-Ee & Hejazi, 2018). It is important to recognize that the field of educational materials is interdisciplinary, with a complex development and review process tailored to various academic areas and goals. Hence, results from a single perspective alone may lack persuasiveness and broad generalizability. Regarding empirical research methods, quantitative research dominates, while qualitative methods are not widely employed; there is a scarcity of cross-sectional comparative studies, as well as fewer longitudinal studies.

Thirdly, the research content lacks systematicity. The existing theoretical literature is predominantly macroscopic and discursive, and the integration of theory and practice falls short in serving actual English teaching. On one hand, the existing theoretical literature is mostly macro-discursive, with only a handful of systematic studies combining theory and practice to effectively support English teaching. On the other hand, most empirical studies concentrate on the characteristics of textbook writing and analysis of text content but overlook the interactions and changes among teachers, students, and teaching contexts during the process of using textbooks. Generally, most of the research focus on English teaching materials in higher education and middle and secondary education. The research to date has not focused on English teaching materials at the lower or basic education level.

This study is constrained by its exclusive focus on literature related to textbook evaluation in the past decade. Consequently, this SLR might have omitted relevant articles published before 2014. Additionally, given the authors' subjective decision-making and knowledge limitations, errors may not have been entirely avoided.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

References

Aghazadeh, Z. (2015). An Evaluation of High School English Textbooks in Iranian EFL Context: Teachers' versus Learners' Perceptions. *Journal of Language Teaching and Research*, *6*(5), 1115. https://doi.org/10.17507/jltr.0605.26

Ahmed, F. A. A., Mohammadzadeh, B., & Mazlum, F. (2023). An in-depth analysis of the representation of speech acts and language functions in Libyan public high school English textbooks. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1056745

Ahour, T., Towhidiyan, B., & Saeidi, M. (2014). The Evaluation of "English Textbook 2" Taught in Iranian High Schools from Teachers' Perspectives. *English Language Teaching, 7*(3), 150. https://doi.org/10.5539/elt.v7n3p150

Alharbi, A. T. (2022). Mega Goals: Evaluating the textbooks for language learning strategies content. *International Journal of Early Childhood Special Education*, *14*(3), 382-393.

Amerian, M., & Tajabadi, A. (2020). The role of culture in foreign language teaching textbooks: An evaluation of New Headway series from an intercultural perspective. *Intercultural Education*, *31*(6), 623-644. https://doi.org/10.1080/14675986.2020.1747291

Amiri, A., & Rezvani, R. (2021). A tale of three official English textbooks: An evaluation of their horizontal and vertical alignments. جستارهای زبانی, *12*(3), 51-79.

Arshad, A., Shah, S. K., & Ahmad, M. (2020). Investigating cultural contents in English language teaching materials through textbook evaluation. *Journal of Language and Cultural Education*, *8*(2), 127-145. https://doi.org/10.2478/jolace-2020-0017

Asiyaban, A. R., & Zamanian, M. (2014). Scrutinizing the Appropriateness of the Intermediate ILI English Series in the EFL Context of Iran. *Theory and Practice in Language Studies*, *4*(6), 1257-1265. https://doi.org/10.4304/tpls.4.6.1257-1265

Azarfam, A. A. Y., & Noordin, N. (2014). Evaluating an English Textbook for Application in Iranian EFL Academic Context. *Theory and Practice in Language Studies*, *4*(3), 618-623. https://doi.org/10.4304/tpls.4.3.618-623

Bababayli, I., & Kiziltan, N. (2020). A comparative sociopragmatic analysis of the dialogues in Turkish and Azerbaijani B1-B2 EFL textbooks. *Dil ve Dilbilimi Çalışmaları Dergisi*, *16*(3), 1500-1522. https://doi.org/10.17263/jlls.803869

Banaruee, H., Farsani, D., & Khatin-Zadeh, O. (2023). Culture in English Language Teaching: A curricular evaluation of English textbooks for foreign language learners. *Front. Educ., 8,* 1012786.

Benabed, A., & Mehdaoui, A. (2023). Algerian Third Year English Coursebook Scrutiny: 'My Book of English' as a Sample. *LLT Journal: A Journal on Language and Language Teaching*, *26*(2), 711-731. https://doi.org/10.24071/llt.v26i2.6884

Bouzid, H. A. (2017). An Evaluation of Selected Moroccan ELT Textbooks: A Standards-based Approach perspective. *Indonesian Journal of Applied Linguistics*, *7*(1), 229. https://doi.org/10.17509/ijal.v7i1.6879

Bui, L. T. N., & Nguyen, M. T. T. (2023). A Pragmatic Analysis of Vietnamese EFL Textbooks: The Case of Disagreement. *Journal of Asia TEFL*, *20*(4), 754-772. https://doi.org/10.18823/asiatefl.2023.20.4.1.754

Cao, D. T. P., Pho, P. D., & Dang, N. A. C. (2022). The Treatment of Collocations in English Textbooks for Vietnamese students. *TEFLIN Journal - A Publication on the Teaching and Learning of English*, *33*(2), 220. https://doi.org/10.15639/teflinjournal.v33i2/220-239

Cunningsworth, A. (1995). *Choosing your coursebook*. Macmillan.

Dennis, M. O. (2018). A Comparison of Readability and Understandability in Second Language Acquisition Textbooks for Pre-service EFL Teachers. *The Journal of AsiaTEFL*, *15*(3), 750-765. https://doi.org/10.18823/asiatefl.2018.15.3.12.750

Derakhshan, A. (2021). 'Should textbook images be merely decorative?': Cultural representations in the Iranian EFL national textbook from the semiotic approach perspective. *Language Teaching Research*, *28*(1), 79-113. https://doi.org/10.1177/1362168821992264

Goodarzi, A., Weisi, H., & Yousofi, N. (2021). Learners on the Move: A Mixed Methods Evaluation of Iranian EFL Textbooks. *MEXTESOL Journal*, *45*(4), n4. https://doi.org/10.61871/mj.v45n4-10

Haddaway, N. R., Macura, B., Whaley, P., & Pullin, A. S. (2018). ROSES Reporting standards for Systematic Evidence Syntheses: pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environmental Evidence*, *7*, 1-8. https://doi.org/10.1186/s13750-018-0121-7

Haghighi, S. (2014). Evaluation of "First Certificate Gold Course Book": Evidence from Students of a Private English Institute in Iran. *International Journal of Instruction*, *7*(2), 149-164.

Hamidi, H., Bagheri, M., Sarinavaee, M., & Seyyedpour, A. (2016). Evaluation of Two General English Textbooks: New Interchange 2 vs. Four Corners 3. *Journal of Language Teaching and Research*, *7*(2), 345. https://doi.org/10.17507/jltr.0702.13

Hessamy, G., & Mohebi, M. (2014). Perceptions of Medical Students and EFL Instructors of Their EAP Materials, Challenges and Implications for Iranian EAP Instructors. *Journal of Language Teaching and Research*, *5*(1), 183-192. https://doi.org/10.4304/jltr.5.1.183-192

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., ... Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for information*, *34*(4), 285-291. https://doi.org/10.3233/EFI-180221

Hooman, Y. (2014). The Evaluation of Iranian EFL Textbooks from Post Method Principles Pedagogy. *Theory and Practice in Language Studies*, *4*(5), 1026-1032. https://doi.org/10.4304/tpls.4.5.1026-1032

Karamifar, M., Barati, H., & Youhanaee, M. (2014). Local Evaluation Criteria for Global Textbooks: A Case Study from Iran. *Theory and Practice in Language Studies*, *4*(5), 923-930. https://doi.org/10.4304/tpls.4.5.923-930

Kraus, S., Breier, M., & Das í Rodr íguez, S. (2020). The art of crafting a systematic literature review in entrepreneurship research. *International Entrepreneurship and Management Journal*, *16*(3), 1023-1042. https://doi.org/10.1007/s11365-020-00635-4

Li, X., & Deocampo, M. F. (2021). An empirical evaluation of two college English textbooks in China based on the China Standards English Framework. *Arab World English Journal*, *12*(3), 323-337. https://doi.org/10.24093/awej/vol12no3.22

Lockwood, C., Munn, Z., & Porritt, K. (2015). Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. *JBI Evidence Implementation*, *13*(3), 179-187. https://doi.org/10.1097/XEB.0000000000000062

Maleki, A., Mollaee, F., & Khosravi, R. (2014). A Content Evaluation of Iranian Pre-university ELT Textbook. *Theory and Practice in Language Studies*, *4*(5), 995-1000. https://doi.org/10.4304/tpls.4.5.995-1000

Manoochehr, J. (2013). Evaluation of Two Popular EFL Coursebooks. *International Journal of Applied Linguistics & English Literature*, *2*(6), 194-201. https://doi.org/10.7575/aiac.ijalel.v.2n.6p.194

Morales, B. C. (2019). Readability and types of questions in Chilean EFL high school textbooks. *TESOL Journal*, *11*(2), 1-15. https://doi.org/10.1002/tesj.498

Namaghi, S. A. O., Moghaddam, M. R. S., & Tajzad, M. (2014). Theorizing Teachers' Perspectives on an EFL Textbook for Public High Schools of Iran: A Grounded Theory. *English Language Teaching*, *7*(10), 123. https://doi.org/10.5539/elt.v7n10p123

Nazim, M. (2021). Textbook Evaluation: A Case Study of Cutting Edge. *Journal of Language Teaching and Research*, *12*(6), 1021-1033. https://doi.org/10.17507/jltr.1206.18

Nu, A. T. T., & Murray, J. (2020). Pragmatic content in EFL textbooks: an investigation into Vietnamese national teaching materials. *Teaching English as a Second or Foreign Language--TESL-EJ*, *24*(3), 1-28. Retrieved from https://files.eric.ed.gov/fulltext/EJ1275813.pdf

Orfan, S. N., Noori, A. Q., & Akramy, S. A. (2021). Afghan EFL instructors' perceptions of English textbooks. *Heliyon*, *7*(11), e08340. https://doi.org/10.1016/j.heliyon.2021.e08340

P érez-Hern ández, L. (2019). From research to the textbook: Assessing speech acts representation in course book series for students of English as an L2. *Revista Espa ñola de Ling ü ística Aplicada/Spanish Journal of Applied Linguistics*, *32*(1), 248-276. https://doi.org/10.1075/resla.17024.per

Rangaraju, S. (2020). A Critical Evaluation of Various English Books. *International Journal of Scientific & Technology, Research, 9*(3), 3977-3980. https://doi.org/10.35940/ijitee.A5067.119119

Rashidi, N., & Kehtarfard, R. (2014). A Needs Analysis Approach to the Evaluation of Iranian Third-Grade High School English Textbook. *SAGE Open*, *4*(3), 215824401455170. https://doi.org/10.1177/2158244014551709

Safa, M. A., Moradi, M., & Hamzavi, R. (2015). Iranian EFL Teachers and Learners Perspective on Potentiality of Top Notch Series for Intercultural Competence Development. *Iranian Journal of Language Teaching Research*, *3*(2), 47-66.

Santos, L. M. D. (2020). The Application of a Textbook Evaluation Checklist: A Research Study of English as a Foreign Language Textbook. *Universal Journal of Educational Research*, *8*(9), 3864-3872. https://doi.org/10.13189/ujer.2020.080910

Shafril, H. A. M., Krauss, S. E., & Samsuddin, S. F. (2018). A systematic review on Asian's farmers' adaptation practices towards climate change. *Sci Total Environ*, *644*, 683-695. https://doi.org/10.1016/j.scitotenv.2018.06.349

Shak, M. S. Y., Albakri, I. S. M. A., Shukor, S. S., & Tahir, M. H. M. (2022). Criteria in English language textbook evaluation checklists: A systematic review. *IARTEM e-journal*, *14*(1), 1-19.

Sheldon, L. E. (1988). Evaluating ELT textbooks and materials. *ELT Journal*, *42*(4), 237-246. https://doi.org/10.1093/elt/42.4.237

Shi, H. (2022). A metaphor analysis of EFL graduate students' beliefs about an EAP textbook. *Frontiers in Education*, *7, 972996*. https://doi.org/10.3389/feduc.2022.972996

Simsek, M. R. (2022). Exploring the Flexibility of ESP Materials Through the IPO Model: Corpus and Consumer Insights from the Turkish EFL Context. *Journal of Teaching English for Specific and Academic Purposes, 111*. https://doi.org/10.22190/JTESAP2201111S

Soto, S. T., & Cevallos, L. F. E. (2021). An Analysis of EFL Textbooks in Public Secondary Schools of Ecuador. *Mextesol Journal*, *45*(4), n4. https://doi.org/10.61871/mj.v45n4-8

Syairofi, A., Mujahid, Z., Mustofa, M., Ubaidillah, M. F., & Namaziandost, E. (2022). Emancipating SLA findings to inform EFL textbooks: A look at Indonesian school English textbooks. *The Asia-Pacific Education Researcher*, 1-12. https://doi.org/10.1007/s40299-022-00642-9

Vahdat, S., Anai Sarab, M. R., & Ghobadi, S. (2020). Evaluating the Cultural Appropriacy of Commercial English Language Teaching Textbooks in the Iranian Context. *Teaching English Language*, *14*(2). https://doi.org/10.22132/tel.2020.119868

Vasheghani, M. (2020). Investigating inside reading textbook series: layout and coherence in focus. *Applied Linguistics Research Journal*, *4*(2), 68-80.

Zare-Ee, A., & Hejazi, S. Y. (2018). Iranian University learners' and Teachers' views on Adopted and Locally-Developed English Language Teaching Textbooks. *International Journal of Instruction*, *11*(3), 291-308. https://doi.org/10.12973/iji.2018.11321a

Appendix A

Research articles included in SLR

Research Title	Author/Year
Algerian Third Year English Coursebook Scrutiny: 'My Book of English' as a Sample	Benabed & Mehdaoui, 2023
Culture in English Language Teaching: A curricular evaluation of English textbooks for foreign language learners	Banaruee et al., 2023
A metaphor analysis of EFL graduate students' beliefs about an EAP textbook	Shi, 2022
The Treatment of Collocations in English Textbooks for Vietnamese Students	Cao et al., 2022
Learners on the Move: A Mixed Methods Evaluation of Iranian EFL Textbooks	Goodarzi et al., 2021
Exploring the Flexibility of ESP Materials Through the IPO Model: Corpus and Consumer Insights From the Turkish EFL Context	Simsek, 2022
Evaluating the Cultural Appropriacy of Commercial English Language Teaching Textbooks in the Iranian Context	Vahdat et al., 2020
'Should textbook images be merely decorative?': Cultural representations in the Iranian EFL national textbook from the semiotic approach perspective	Derakhshan, 2021
Pragmatic Content in EFL Textbooks: An Investigation into Vietnamese National Teaching Materials	Nu & Murray, 2020
A Comparison of Readability and Understandability in Second Language Acquisition Textbooks for Pre-service EFL Teachers	Dennis, 2018
Iranian University Learners' and Teachers' Views on Adopted and Locally-Developed English Language Teaching Textbooks	Zare-Ee & Hejazi, 2018
An Evaluation of Selected Moroccan ELT Textbook: A Standards-based Approach Perspective	Bouzid, 2017
Iranian EFL teachers and learners perspective on potentiality of Top Notch series for intercultural competence development	Safa et al., 2015
An Evaluation of High School English Textbooks in Iranian EFL Context: Teachers' versus Learners' Perceptions	Aghazadeh, 2015
The Evaluation of "English Textbook 2" Taught in Iranian High Schools from Teachers' Perspectives	Ahour et al., 2014
Scrutinizing the Appropriateness of the Intermediate ILI English Series in the EFL Context of Iran	Asiyaban & Zamanian, 2014
A Content Evaluation of Iranian Pre-university ELT Textbook	Maleki et al., 2014
A Needs Analysis Approach to the Evaluation of Iranian Third-Grade High School English Textbook	Rashidi & Kehtarfard, 2014
Perceptions of Medical Students and EFL Instructors of Their EAP Materials, Challenges and Implications for Iranian EAP Instructors	Hessamy & Mohebi, 2014
Evaluation of "*First Certificate Gold Course Book*": Evidence from Students of a Private English Institute in Iran	Haghighi, 2014
Afghan EFL instructors' perceptions of English textbooks	Orfan et al., 2021
A Tale of Three Official English Textbooks: An Evaluation of Their Horizontal and Vertical Alignments	Amiri & Rezvani, 2021
The role of culture in foreign language teaching textbooks: an evaluation of *New Headway* series from an intercultural perspective	Amerian & Tajabadi, 2020
A comparative sociopragmatic analysis of the dialogues in Turkish and Azerbaijani B1-B2 EFL textbooks	Bababayli & Kiziltan, 2020
A Pragmatic Analysis of Vietnamese EFL Textbooks: The Case of Disagreement	Bui & Nguyen, 2023
Mega Goals: Evaluating the textbooks for language learning strategies content	Alharbi, 2022
An Empirical Evaluation of Two College English Textbooks in China Based on the China Standards English Framework	Li & Deocampo, 2021
Investigating Inside Reading Textbook Series: Layout and Coherence in Focus	Vasheghani, 2020
An in-depth analysis of the representation of speech acts and language functions in Libyan public high school English textbooks	Ahmed et al., 2023
Investigating cultural contents in English language teaching materials through textbook evaluation	Arshad et al., 2020

Appendix B

List of Studies by Year

Year	Quantity	Studies
2023	4	Ahmed et al., 2023; Banaruee et al., 2023; Benabed & Mehdaoui, 2023; Bui & Nguyen, 2023
2022	4	Alharbi, 2022; Cao et al., 2022; Shi, 2022; Simsek, 2022
2021	5	Amiri & Rezvani, 2021; Derakhshan, 2021; Goodarzi et al., 2021; Li & Deocampo, 2021; Orfan et al., 2021
2020	6	Amerian & Tajabadi, 2020; Arshad et al., 2020; Bababayli & Kiziltan, 2020; Nu & Murray, 2020; Vahdat et al., 2020; Vasheghani, 2020
2018	2	(Amiri & Rezvani, 2021; Derakhshan, 2021; Goodarzi et al., 2021; Li & Deocampo, 2021; Orfan et al., 2021)
2017	1	Bouzid, 2017;
2015	2	Aghazadeh, 2015; Safa et al., 2015
2014	6	Ahour et al., 2014; Haghighi, 2014; Asiyaban & Zamanian, 2014; Hessamy & Mohebi, 2014; Maleki et al., 2014; Rashidi & Kehtarfard, 2014

Appendix C

List of Studies by Country

Country	Quantity	Studies
Afghan	1	Orfan et al., 2021
Algerian	1	
China	2	Li & Deocampo, 2021; Shi, 2022
Iran	15	Aghazadeh, 2015; Ahour et al., 2014; Amerian & Tajabadi, 2020; Amiri & Rezvani, 2021; Haghighi, 2014; Asiyaban & Zamanian, 2014; Banaruee et al., 2023; Derakhshan, 2021; Goodarzi et al., 2021; Hessamy & Mohebi, 2014; Maleki et al., 2014; Rashidi & Kehtarfard, 2014; Safa et al., 2015; Vahdat et al., 2020; Zare-Ee & Hejazi, 2018
Libya	1	Ahmed et al., 2023
Moroccan	1	Bouzid, 2017
Not mention	1	Vasheghani, 2020
Pakistan	1	Arshad et al., 2020
Saudi	1	Alharbi, 2022
South Korea	1	Dennis, 2018
Turkey	1	Simsek, 2022
Turkey&Azerbaijani	1	Bababayli & Kiziltan, 2020
Vietnam	3	Bui & Nguyen, 2023; Cao et al., 2022; Nu & Murray, 2020

Appendix D

List of Studies by Level of Education

Level of Education	Quantity	Studies
Primary	1	Benabed & Mehdaoui, 2023
Middle & Secondary	15	Aghazadeh, 2015; Ahour et al., 2014; Ahmed et al., 2023; Alharbi, 2022; Amiri & Rezvani, 2021; Arshad et al., 2020; Bababayli & Kiziltan, 2020; Banaruee et al., 2023; Bouzid, 2017; Bui & Nguyen, 2023; Derakhshan, 2021; Goodarzi et al., 2021; Nu & Murray, 2020; Rashidi & Kehtarfard, 2014; Vahdat et al., 2020
Higher Education	11	Amerian & Tajabadi, 2020; Haghighi, 2014; Asiyaban & Zamanian, 2014; Dennis, 2018; Hessamy & Mohebi, 2014; Li & Deocampo, 2021; Orfan et al., 2021; Safa et al., 2015; Shi, 2022; Simsek, 2022; Zare-Ee & Hejazi, 2018
Other	3	Cao et al., 2022;; Maleki et al., 2014; Vasheghani, 2020

Appendix E

List of Studies by Subject

Evaluator	Quantity	Studies
Researcher	9	Ahmed et al., 2023; Alharbi, 2022; Amiri & Rezvani, 2021; Arshad et al., 2020; Bababayli & Kiziltan, 2020; Bouzid, 2017; Bui & Nguyen, 2023; Cao et al., 2022; Nu & Murray, 2020
Researcher & student	1	Li & Deocampo, 2021
Student	5	Dennis, 2018; Goodarzi et al., 2021; Rashidi & Kehtarfard, 2014; Shi, 2022; Simsek, 2022
Teacher	5	Ahour et al., 2014; Amerian & Tajabadi, 2020; Asiyaban & Zamanian, 2014; Banaruee et al., 2023; Orfan et al., 2021
Teacher & expert	2	Benabed & Mehdaoui, 2023; Vasheghani, 2020
Teacher & student	7	Aghazadeh, 2015; Haghighi, 2014; Hessamy & Mohebi, 2014; Maleki et al., 2014; Safa et al., 2015; Vahdat et al., 2020; Zare-Ee & Hejazi, 2018
Teacher, student, & expert	1	Derakhshan, 2021

Appendix F

List of Studies by Sample Size of Subject

Sample size of evaluator	Quantity	Studies
from 1 to 10	1	Alharbi, 2022; Nu & Murray, 2020
from 11 to 100	10	Ahour et al., 2014; Amerian & Tajabadi, 2020; Haghighi, 2014; Asiyaban & Zamanian, 2014; Benabed & Mehdaoui, 2023; Dennis, 2018; Derakhshan, 2021; Orfan et al., 2021; Simsek, 2022; Vasheghani, 2020
above 100	11	Aghazadeh, 2015; Banaruee et al., 2023; Goodarzi et al., 2021; Hessamy & Mohebi, 2014; Li & Deocampo, 2021; Maleki et al., 2014; Rashidi & Kehtarfard, 2014; Safa et al., 2015; Shi, 2022; Vahdat et al., 2020; Zare-Ee & Hejazi, 2018
other	7	Ahmed et al., 2023; Amiri & Rezvani, 2021; Arshad et al., 2020; Bababayli & Kiziltan, 2020; Bouzid, 2017; Bui & Nguyen, 2023; Cao et al., 2022

Appendix G

List of Studies by Type of Methodology

Research methodology	Quantity	Studies
mixed	12	(Aghazadeh, 2015; Ahmed et al., 2023; Amerian & Tajabadi, 2020; Amiri & Rezvani, 2021; Benabed & Mehdaoui, 2023; Bouzid, 2017; Derakhshan, 2021; Goodarzi et al., 2021; Hessamy & Mohebi, 2014; Li & Deocampo, 2021; Nu & Murray, 2020; Simsek, 2022)
qualitative	7	Alharbi, 2022; Arshad et al., 2020; Banaruee et al., 2023; Bui & Nguyen, 2023; Cao et al., 2022; Shi, 2022; Vasheghani, 2020
quantitative	11	Ahour et al., 2014; Haghighi, 2014; Asiyaban & Zamanian, 2014; Bababayli & Kiziltan, 2020; Dennis, 2018; Maleki et al., 2014; Orfan et al., 2021; Rashidi & Kehtarfard, 2014; Safa et al., 2015; Vahdat et al., 2020; Zare-Ee & Hejazi, 2018