Thiri Soe
Graduate School of International Cultural Studies
Tohoku University, Japan

https://orcid.org/0009-0006-6161-696X

Investigating Critical Thinking in ELT Textbooks: A Systematic Literature Review of Textbook Evaluation Studies

Abstract

The pivotal role of critical thinking and its integration in English language education are synthesised through systematic literature reviews and meta-analysis studies in the areas of pedagogical interventions, teaching methods, and assessment methods. However, there is little evidence of systematic literature review on the integration of critical thinking into English language teaching materials such as textbooks. In this study, 41 empirical textbook evaluation studies published between 2010 and 2021 were reviewed from critical thinking perspective. The purpose of this study is to provide an overview of English textbook evaluations on critical thinking integration, to identify critical thinking frameworks applied, and to look for possible research gaps among the studies. An extensive literature search was conducted by applying Xiao and Watson's (2017) eight-step systematic literature review method. From a collection of 41 empirical studies, this review focuses on such four aspects as applied critical thinking frameworks, findings, recommendations, and the language levels and content areas of the textbooks studied. The results revealed that (1) Bloom's Taxonomy and Bloom's Revised Taxonomy were most frequently applied critical thinking frameworks, (2) integration of higher-order thinking skills in textbooks was less common than lower-order thinking skills, (3) textbook assessment occurred primarily at the middle and high school levels more than at other levels, and (4) methodological considerations regarding the reliability and validity of the coding process in textbook content analysis received little attention. This study contributes a synthesised literature background of English textbook evaluation with recommendations for methodological improvements in future studies.

Keywords: systematic literature review, critical thinking, ELT textbook evaluation, higher order thinking skills, lower order thinking skills

Critical Thinking Integration in ELT Textbooks

Mastery of critical thinking skills is a premium goal in education sector. Success in a profession is directly related to a person's superior critical thinking skills (Cottrell, 2017). In the age of technology where information is readily available at one mouse's click, educational goals are directed from rote learning toward enhancing transferable skills like critical thinking across subject specific contexts (Puig et al., 2019). As such, teachers, researchers, and education specialists have focused on improving learners' critical thinking skills through specialised course content that incorporates critical thinking activities. Critical thinking across curricula in subject-specific contexts is highly recommended for the development of learners in accordance with educational objectives (Ennis, 2018). Therefore, materials and activities used in language classrooms have become key factors in enhancing language learners' critical thinking skills (Shirkhani & Fahim, 2011). A teacher's choice of teaching methods and materials also directly affects thinking practices and training students receive (Thompson, 2011). Examining the degree of emphasis on critical thinking in ELT textbooks uncovers potential contribution of the textbooks to language learning (Birjandi & Alizadeh, 2013; Solihati & Hikmat, 2018). Thus, textbook analysts and pedagogues of English language teaching are increasingly examining the integration of critical thinking skills in language teaching materials. Despite the growing interest in critical thinking skills and activities developed in ELT materials, there is a lack of systematic literature review in ELT textbook evaluation studies from critical thinking perspective. Research syntheses are necessary since they summarise the breadth and depth of the existing literature and can also indicate research gaps that raise new research questions for further methodologically and theoretically improved studies. For these reasons, this systematic literature review study attempts to survey the existing literature on the analysis and evaluation of ELT textbooks critical thinking integration perspective.

Literature Review

Critical thinking is known to have a variety of proposed definitions. In the course of its evolution, scholars, philosophers, psychologists, and educationalists have each attempted to define the concept of critical thinking skills in their own way, and there are overlapping characteristics among these definitions (Nilson, 2021). There is neither a universal agreed-upon set of critical think-

ing skills nor an educational approach for implementing them. The scope and depth of critical thinking in literature is so massive that it cannot be easily grasped as a specific idea that can be applied in any discipline in any context. In 2013, Moore conducted a qualitative study in three academic disciplines such as philosophy, history, and literary and cultural studies to define the term critical thinking as the basis for understanding and application by scholars. These disciplines were intentionally chosen because of their close relationship to education. As a result of the interview responses, Moore concluded four main concepts of critical thinking. These concepts are (1) the ability to judge ideas, (2) a skeptical and provisional view of knowledge, (3) originality of ideas or modest contribution to a body of knowledge, and (4) careful and sensitive reading of texts or input of information. In addition, other concepts peripheral to critical thinking included rational thinking, adopting an ethical and activist stance, and self-reflection.

In 2018, Ennis, a prominent figure in the field of critical thinking education, said that scholars in different disciplines define critical thinking differently depending on their stance in their field of expertise. However, as a way to cut the same pie from different angles, their definitions and assumptions are all broadly similar. Ennis clarified the concept of critical thinking by summarising all of his research over the past 30 years in one expression as, "rational reflective thinking focused on deciding what to believe and do (p. 166)." In an attempt of helping students develop their critical thinking skills, Halpern (1999, p. 70) elaborated on the concept of critical thinking as follows.

> Critical thinking refers to the use of cognitive skills or strategies that increase the probability of a desirable outcome. Critical thinking is purposeful, reasoned, and goal-directed. It is the kind of thinking involved in solving problems, formulating inferences, calculating likelihoods, and making decisions.

Halpern's core definition of critical thinking skills includes five key elements: (1) reasoning skills, (2) hypothesis testing, (3) analysis of arguments, (4) analysis of possibilities and uncertainties, and (5) decision making and problem solving. As discussed so far, several characteristics of the concept of critical thinking have been proposed by different authors from different disciplines. Thus, the practice of critical thinking in the process of language acquisition implies a rational and open-minded acceptance of language input and output. Moreover, effective communication and critical thinking skills are rated as the most important intellectual skills demanded by employers in the 21st century job industry. In an attempt to enquire into the interpretation of the meaning of critical thinking and language competency, Jones et al. (1995) surveyed a total of 600 teachers, employers, and policy makers about their definitions

of writing, speaking, listening, and critical thinking and their expectations for how college graduates should develop these skills to become effective employees at work and informed citizens in society. The survey result stated that for language input, learners should be able to identify biased or misleading information. Learners require the ability to see through indirect persuasion, information and facts that are irrelevant to an argument, and information and questions that evoke speakers' preferred responses. For language output through writing, learners should have knowledge of audience awareness (considering potential readers of their writing), purposeful writing, and text organisation. For language output through speaking, learners should be able to exchange information appropriately, manage conversations and group communication. Behar-Horenstein and Niu (2011) also supported this idea by explaining that critical thinkers must not only examine and evaluate information beforehand in their critical thinking processes, but must also take responsibility for challenging and monitoring their own thinking and arguments. These key concepts of critical thinking in the language learning aspect are also consistent with Moore and Parker (2009), who delved deeply into the concept of critical thinking and developed a long list of 21 characteristics as a definition of critical thinking. According to their definition, which summarises these 21 characteristics, critical thinking can be divided into two main categories: using logical thinking to evaluate the credibility of information and using logical thinking to construct one's arguments. Therefore, we posit that from the domain of the language learning, the concept of critical thinking skill can be boiled down into two aspects: language learners' application of rational thinking in language input and output processing, in other words, receptive and productive language processing.

The Roles of Critical Thinking in English Language Education

The role of critical thinking is essential in English language learning for a number of reasons. Increasing access to the internet and information through digital media reinforces the urgent need for learners' critical thinking development. Dummett and Hughes (2019) posited that through critical thinking activities, language learners acquire four different types of literacy: information literacy, media literacy, visual literacy, and intercultural literacy, all of which foster learners' effective and constructive communication. For authentic communication, it is not enough for learners to memorise and repeat chunks of a given language examples. To communicate effectively through the target language, learners must analytically investigate given information, personalise it, and solve problems, all of which require critical thinking skills (Hughes, 2014). Through critical thinking integrated activities and tasks in language classes, learners are able to perceive and process information critically. Long-

term memory is also enhanced when learners deeply and critically explore language input through critical thinking activities (Dummett, 2016). Learners' learning memories are categorised as implicit and explicit memories, also called procedural and declarative memories, respectively. In first language acquisition, learners acquire their native language through implicit memory, which is unconscious and instinctive in nature. In foreign language acquisition, however, learners acquire the target language through explicit memory, or conscious learning. Therefore, by encouraging learners to engage in critical thinking tasks and practices over time, they move from a deep critical exploration of language input to a more instinctive sense of the target language being experienced. As a result, critical thinking activities reinforce learners' long-term memory in the target language acquisition process and move learners toward instinctive and implicit memory (Dummett & Hughes, 2019).

Not only are critical thinking skills and English language proficiency parallel skills needed to meet the demands of the 21st century employment industry, but there is a strong positive correlation between the two. According to the literature on the role of critical thinking in ELT education, pedagogical approaches for language learning that incorporate critical thinking have been widely discussed, and its role continues to expand in English language teaching programs. Integration of critical thinking in English language teaching is established in two areas of interest: language teacher's instructional methods that incorporate critical thinking and development of language teaching materials that reinforce critical thinking (Shirkhani & Fahim, 2011). In the aspect of instructional approaches, specialists and researchers provided teaching instructions and methods to help learners engage in critical thinking fostering activities in language classrooms. For example, Hughes (2014) presented a list of 20 ELT classroom activities and demonstrated how different problem solving activities can be integrated at different language proficiency levels. To help college students develop the ability to evaluate the reliability of information intake, Halpern (2002) set up a four-part critical thinking instructional model consisting of critical thinking dispositions, critical thinking skills approach, structural training, and metacognitive monitoring. Alnofaie (2013) proposed a framework for incorporating critical thinking activities into four language skills. Thus, English language teachers have become aware of effective teaching-learning techniques that enhance language learners' thinking skills and language skills simultaneously. English language teachers' classroom experiments and research findings provide guidance and effective recommendations on classroom activities, and areas that teachers should focus on to enhance learning opportunities. Notable examples include Lin's (2018) critical thinking infusion approach in Chinese high school learners' English composition, which reportedly improved both critical thinking skills and writing performance. Golpour (2014) found that Iranian EFL learners with high critical thinking

skills showed significant performance in various writing modes, including expository and argumentative writing. In Yang and Gamble's (2013) classroom experiment using critical thinking-enhanced activities such as debate and peer critique, the experimental group made significant gains in reading and listening proficiency compared to the control group that received no critical thinking integration instruction. Fahim et al. (2012) found in their study that critical thinking training had a facilitative effect on reading performance of high and low proficiency EFL students. Sanavi and Tarihat (2014) showed that explicitly introducing critical thinking skills improved learners' speaking skills compared to learners who did not receive critical thinking reinforcement.

From the perspective of language instruction, as discussed above, experimental studies in English classrooms have emphasised that incorporating critical thinking skills into classroom interventions improves language proficiency in all four skills. However, the development of language learners' critical thinking from the area of language teaching materials has not yet been well explored, especially in terms of a systematic literature review. Furthermore, there are still no fixed parameters or predominantly established working models applied to the evaluation of language materials. Therefore, the purposes of this study are to scrutinize the existing literature on ELT textbook evaluation research from critical thinking perspective, to identify widely used critical thinking frameworks in the evaluation, and to identify possible research gaps in the area.

Evaluating Critical Thinking Incorporation in ELT Textbooks

Throughout our discussion of critical thinking in language teaching materials, we refer interchangeably to ELT course books, ELT textbooks, English language textbooks, and English language teaching materials as the same concept in the context of this study. We also used the two terms of *textbook evaluation* and *textbook analysis* interchangeably for examining and evaluating textbooks' contents. Research on the evaluation of English language teaching materials can be summarised into two dimensions: (1) how to evaluate and (2) what to evaluate. There are three main stages of evaluation methods for teaching materials: pre-use evaluation, in-use evaluation, and post-use evaluation (McGrath, 2002). These stages differ with respect to the time they are conducted and the objectives of the evaluation. Pre-use evaluation is conducted in selecting teaching materials before a tentative textbook is actually determined for use. Evaluators like teachers and concerned professionals observe the potential that their tentatively chosen textbooks can promise. In the in-use evaluation stage, evaluators actually use the textbook while observing whether it fits the actual classroom situation and meets its claimed potential. The post-use evaluation phase is conducted after a textbook has been in use for a period

of time to help teachers figure out what of the possibilities the textbook initially promised actually worked and what did not. This third stage can involve learners' performance assessment. It is the stage where teachers reflect upon the difficulties and gaps that need to be supplemented and adapted to learners' needs and learning situations. Ellis (1997) called pre-use evaluation *a predictive evaluation* (examining course materials before using) and post-use evaluation as a *retrospective evaluation* (examining materials after using). For the predictive evaluation, language teachers either evaluate textbooks by themselves or rely on other expert reviewers. Post-use or retrospective evaluation poses some challenges and difficulties for teachers to carry out because they are time-consuming as empirical case studies.

In this section, we will only illuminate the fact that there is no place yet for critical thinking in the evaluative criteria established so far in the literature. We cannot cover the complete literature background regarding the criteria and parameters used for ELT textbook evaluation. Here are a few examples among the huge number of ELT textbook evaluation parameters or checklists theorised by experts in the field. Williams (1983) proposed an evaluative scheme that consists of four criteria: up-to-date methodology, guidance for non-native teachers, meeting the learning goals of language learners, and relevance to the socio-cultural environment. Sheldon (1988) proposed a 17-item-list of common core factors for ELT textbook assessment, which mainly focuses on practicability, flexibility, accessibility, and layout of teaching materials. McDonough et al. (2013) proposed external and internal evaluations for the analysis of textbook contents. External evaluation looks into the contents and organisation of textbooks and other general factors such as availability of teachers' books, cultural and gender biases, layout and presentation, etc. If the results of the external evaluation indicate that the subject material is appropriate, the evaluation proceeds to internal evaluation, where the sequence of materials and exercises, the relationship between exercises and tests, and the suitability of the exercises and texts for the learning style are examined in detail. In 2015, Brown and Lee proposed a list of criteria to evaluate an ELT textbook. Their criteria are focused on the perspectives of program/course, teaching approaches, language skills, the practicality of the materials, and the availability of the supplementary materials. These aforementioned examples show how evaluation criteria all focus on the authenticity and relatedness between the teaching materials and learners' real life experiences. In other words, they mean that appropriate tasks and activities that activate learners' engagement and practical application to real life situations are highly valued in ELT material development. Learners' language practices and language input should be closely related to real-life problem solving and experiences that require learners to apply critical thinking.

Even though researchers and experts in the field of materials development have developed a rich literature on evaluation criteria for language materials,

their methods and applications are specific to each language learning situation and cannot be transferred directly to other situations without modifications (Tomlinson, 2012). However, there is one basic principle that is common to all learning situations and to all learning materials developed in different teaching and learning situations. The principle is that applied language instructions and language teaching materials should meet the actual needs and goals of a particular group of language learners in a particular context (McDonough et al., 2013). As supported by classroom experiments and case studies on language teaching methods that incorporate critical thinking, the development of critical thinking skills is essential in the process of language acquisition, regardless of different learning goals or learning situations. Still, there are currently no parameters that explicitly target the inclusion of critical thinking activities in ELT textbooks. Littlejohn (2011) also emphasised the need for evaluation criteria that critically question about the status of learning autonomy, engagement in problem-solving tasks, and emphasis on learner-centred approaches. Prior to the development of such assessment criteria, research on the integration of critical thinking into ELT curriculum is a foundation step in a future roadmap for developing assessment criteria for ELT textbooks from critical thinking perspective.

Materials and Methods

In our systematic literature review we applied Xiao and Watson's (2017) eight-step approach. Figure 1 depicts the flow of the eight-step review process, including key decisions and protocols made based on the availability of relevant literature and the purpose of the review.
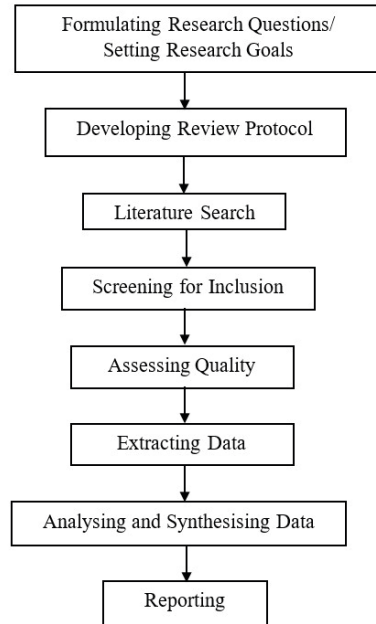
Step 1: Formulating Research Questions or Setting Research Goals

Critical thinking plays a prominent role in English language education worldwide. In addition, evaluation criteria for ELT textbooks have long been developed to enhance the quality and sustainability of the teaching materials. However, evaluating ELT textbooks and learning materials from critical thinking integration perspective is relatively new in its field. Consequently, there are no summarised and organised literature review studies on the topic so far.

Figure 1

Flow of the Systematic Literature Review Process Adapted from Xiao and Watson (2017)

```
┌─────────────────────────────┐
│ Formulating Research Questions/ │
│    Setting Research Goals      │
└─────────────────────────────┘
              │
              ▼
    ┌─────────────────────┐
    │ Developing Review Protocol │
    └─────────────────────┘
              │
              ▼
      ┌─────────────────┐
      │ Literature Search │
      └─────────────────┘
              │
              ▼
    ┌─────────────────────┐
    │ Screening for Inclusion │
    └─────────────────────┘
              │
              ▼
      ┌─────────────────┐
      │ Assessing Quality │
      └─────────────────┘
              │
              ▼
      ┌─────────────────┐
      │ Extracting Data   │
      └─────────────────┘
              │
              ▼
  ┌───────────────────────────┐
  │ Analysing and Synthesising Data │
  └───────────────────────────┘
              │
              ▼
        ┌───────────┐
        │ Reporting │
        └───────────┘
```

It is worth examining the depth and scope of the existing literature, research findings, and applied critical thinking frameworks and methods in EFL textbook evaluation. In order to scrutinise the current state of ELT textbook evaluation studies from critical thinking integration aspect and research trends, and to identify possible research gaps in this domain of study, a systematic review of previous literature is necessary. For these reasons and objectives, this study is conducted by critically synthesising the previous empirical studies within the last eleven years. The parameters to be observed were determined by referring to Cooper's (1988) *focus* of the literature review. According to Cooper, the criterion called *focus* encompasses such factors as (1) research outcomes, (2) research methods, (3) theories, and (4) practices or applications. With reference to these four characteristics for the focus of literature review, we set the research objectives to examine the following five points in our textbook evaluation study:

– critical thinking frameworks used in the ELT textbook evaluation studies;
– results of the studies on the percentage of critical thinking integrated tasks in textbooks;
– levels and language areas of ELT textbooks being examined;
– validity and reliability of the studies, and;
– recommendations made by the studies.

Step 2: Developing Review Protocol

This is the planning stage for the literature review process, including re-search objectives, questions, search strategies, inclusion criteria, data extraction, analysis, synthesis and interpretation, and writing a report. All of these will appear in a sequential process in the following stages. In this stage, limiting the scope of the study and inclusion criteria are the key decisions because they are the roadmaps of the study. The following three criteria are review protocols developed in order to limit the scope of the study:
1. The retrieved studies are conducted in EFL context.
2. The study period is from 2010 to 2021 (12 years).
3. The retrieved studies are conducted, focusing on critical thinking manifestation.

Step 3: Literature Search

In the literature search, we used a purposive sample approach with keyword searching, forward searching, and backward searching, targeting the studies that focused on ELT textbook evaluation in EFL context. Databases used for literature search include Web of Science, ERIC, Elsevier, Scopus, Google Scholar, Science Direct, and PsychInfo. Keywords used in searching are English language textbook evaluation and critical thinking, critical thinking in language teaching materials, and critical thinking in ELT/EFL textbooks.

Step 4: Screening for Inclusion

Step 4, the screening process is the preliminary screening stage of Step 5, which is the quality assessment. In this Step 4, the retrieved literature tempo-rarily saved in a file was screened to determine whether or not to include it in the review. That appraisal was done by reviewing the abstracts of the studies and skimming through the entire article. Studies that did not follow the review protocol were excluded.

Step 5: Assessing Quality

This step is a detailed evaluation of the content. The recruited studies were screened in more detail by reading the full texts of the studies. The decision to include the studies was made in reference to three criteria developed in the review protocol step. It is worth mentioning here that the inclusion of the studies was not restricted to peer-reviewed articles due to the little availability of the

studies and their role in EFL context. After excluding studies not relevant to the purpose and scope of the review and after removing two duplicate publications, a total of 41 empirical studies were obtained for data extraction, analysis, and interpretation.

Step 6 and Step 7: Extracting Data, Analysing and Synthesising

Data extraction, analysing, and synthesising were performed, following Lune and Berg's (2017) three concurrent flows method: (1) data reduction, (2) data display, and (3) conclusion and verification. First, we developed coding themes or analytic categories based on the review protocol. Second, raw data were coded and transformed into interpretable themes. The raw data were first recorded in a Word file and then converted to an Excel sheet, organised by category to be analysed, and frequencies were counted. Data display is an inseparable process of data reduction. In that step, data were displayed in visualised forms for the analysis such as tables, figures, excel sheets, and tally sheets. From these visualisations, interrelated themes and patterns were discovered. Third, in the conclusion and verification step, the analytic conclusion was made through evaluation and decision making that occurred throughout the analysis. That analytic conclusion was verified by retracing the analytic steps of the data by the same researcher sometime after the first analysis had been done.

Step 8: Reporting

The final stage is reporting the whole literature review work to the academic realm to disseminate the study such as this academic paper to be engaged with the concerned scholars in the field. This systematic literature review is reported following the PRISMA statement (Moher et al., 2009), a comprehensive checklist for reporting a systematic literature review.

Findings and Discussion

This section presents the findings of the data derived from the systematic literature review that examines five aspects from the previous empirical studies such as (1) critical thinking frameworks used in the ELT textbook evaluation studies, (2) results of the primary studies on the percentage of critical thinking

tasks in textbooks, (3) language levels and language areas of ELT textbooks being examined, (4) validity and reliability of the studies, and (5) recommendations made by the studies.

Critical Thinking Frameworks Applied in the Studies

The first central observation, as given in Table 1, is the application of critical thinking frameworks in the studies. There are eleven critical thinking frameworks applied by different numbers of studies: 15 studies applied Bloom's Taxonomy (Bloom et al., 1956) and 20 studies applied Bloom's Revised Taxonomy (Anderson et al., 2001) while the rest of the studies applied other critical thinking frameworks such as Facione's Critical Thinking Model (2011), Paul-Elder's Critical Thinking Model (Elder & Paul, 2012), Peterson's Critical Thinking Model (Peterson, 2008), Ilyas's Critical Thinking framework (2015), the survey method of questionnaires and interviews, and *Inference* and *Deduction* as essential sub-thinking skills. It should be noted that four studies are found to be applying other critical thinking frameworks while they were using Bloom's Taxonomy and Bloom's Revised Taxonomy in their analyses. For example, Akrong et al. (2021) used Bloom's Revised Taxonomy and Cummins's (1999) framework: cognitive, academic, and language dimensions. Shuyi and Renandya (2019) used Bloom's Revised Taxonomy and Webb's (2002) Domain of Knowledge Model, and Tabari and Tabari (2015) used Bloom's Taxonomy and Multiple Intelligence Model (Gardner, 1993, 1999). The actual number of reviewed studies is 41, but three of them used two critical thinking frameworks simultaneously. Thus, the total number of studies indicating the use of these all critical thinking frameworks is 44, as shown in Table 1. It should also be noted that there is an extreme imbalance in the frequency of the use of critical thinking frameworks. Each of the critical thinking frameworks other than Bloom's Taxonomy and Bloom's Revised Taxonomy was applied only once. The aim of the present literature review is not to discuss these less frequently applied frameworks. Therefore, in order to explore the central patterns and themes that emerge from the findings, we focused only on the studies that belong to the most frequently applied critical thinking frameworks, that is, Bloom's Taxonomy and Bloom's Revised Taxonomy. As a result, subsequent data extraction and analysis were conducted in studies using these two frameworks, the results of which are presented in the following section.

Table 1

Critical Thinking Frameworks Applied as Evaluation Checklists

Critical Thinking Frameworks	Frequency of Use
Bloom's Revised Taxonomy (2001)	20
Bloom's Taxonomy (1956)	15
Facione's Critical Thinking Model (2011)	1
Paul-Elder's Critical Thinking Model (2012)	1
Peterson's Critical Thinking Model (2008)	1
Ilyas's Critical Thinking Framework (2015)	1
Cummins's framework: cognitive, academic, and language dimensions (1999)	1
Webb's Domain of Knowledge Model (2002)	1
Gardner's Multiple Intelligence Model (1993; 1999)	1
Questionnaire and Interviews	1
Inference and Deduction	1
Total	44

The Most Frequently Applied Critical Thinking Frameworks

In seeking the critical thinking integration inside the textbooks, we limited our focus to the studies that used the two most common critical thinking frameworks, Bloom's Taxonomy and Bloom's Revised Taxonomy. It is not because the other critical thinking models and theories are less important and less credible, but because our study aims to draw a conclusive summary out of possible prominent themes and patterns found in the analyses. Therefore, we synthesised the results of 35 studies using these two taxonomies and summed the percentages obtained from each study. Another good reason to focus only on the two taxonomies is that they are well-established theoretical frameworks for curriculum development and learning goals in the education sector. Bloom's Taxonomy and Bloom's Revised Taxonomy were most frequently used critical thinking frameworks, with Bloom's Revised Taxonomy occupying 20 studies (49%) out of the total reviewed studies, Bloom's Taxonomy occupying 15 studies (36%) out of the total reviewed studies, resulting in a total of 49% of the studies. Figure 2 is the representation of the two taxonomies and overall percentages yielded after adding up the results of critical thinking skills in Bloom's Taxonomy and Bloom's Revised Taxonomy respectively. There are changes in the Revised Taxonomy from the original one; however, the clear classification between lower and higher order thinking skills, shown by a dotted line in Figure 2, remains the same in both taxonomies. Consequently, the studies

came up with the common categorisation of lower and higher order thinking skills in their findings. All 35 studies reached the common conclusion that the ratio of lower-order to higher-order thinking skills is unbalanced. Studies using Bloom's Taxonomy found that 28% of the content of the textbooks analysed activated higher-order thinking skills, with the remaining 72% being lower-order thinking skills. Studies using Bloom's Revised Taxonomy found that 16% of the content of the textbooks analysed activated higher-order thinking skills, with the remaining 84% being lower-order thinking skills. This situation is also emphasized in Krathwohl (2002), where less focus is placed on the higher order thinking skills of the six criteria, and the focus of educational practices and exercises is generally on the lower order thinking level such as knowledge and understanding.

When calculating the overall average percentage of critical thinking skills found within the textbooks, we had to omit some studies that did not analyse and present the integration of critical thinking activities by counting the number of times critical thinking activities were enhanced. For example, Birjandi and Alizadeh (2013) examined the integration of critical thinking skills on a Likert scale and thus we could not examine the average percentage from it. Therefore, this study, which evaluated textbooks using Bloom's Taxonomy, was omitted from the calculation of the overall percentage of all studies using Bloom's Taxonomy. Similarly, Jebbour (2019) did not provide a quantitative calculation of critical thinking activities in the textbooks they examined. The authors simply described which critical thinking skills were tapped into the textbooks and which were not. The reader cannot know what percentage of critical thinking activities were included from these studies. In addition, studies such as Razmjoo and Kazempourfard (2012) were not included in calculating averages of each thinking skill because the thinking skills are only presented cumulatively in two groups, the lower order thinking skills group and the higher order thinking skills group. Such cases occurred in two of the 15 studies using Bloom's Taxonomy and two of the 20 studies using Bloom's Revised Taxonomy. However, we assume that these studies were negligible in number compared to the overall number of studies and were unlikely to affect the overall percentage result. The average percentage of each thinking skill in both taxonomies was calculated from 31 studies (excluding four studies irrelevant to the overall calculation), and the results are presented in Figure 2.

For the information of the readers of this study, a brief historical background of these frequently applied taxonomies should be provided. The fundamental frameworks of the two taxonomies are based on the concept that the attainment of educational goals starts from simple to complex skills as in a hierarchical order of six learning skills in the pyramid structure. The first taxonomy, called Bloom's Taxonomy, was developed by Bloom and his colleagues, Engelhart, Furst, Hill, and Krathwohl from education and psychological domains.

Figure 2

The Two Most Often Referenced Critical Thinking Frameworks: Bloom's Taxonomy and Bloom's Revised Taxonomy

Bloom's Taxonomy (Bloom et al., 1956)   Bloom's Revised Taxonomy (Anderson et al., 2001)
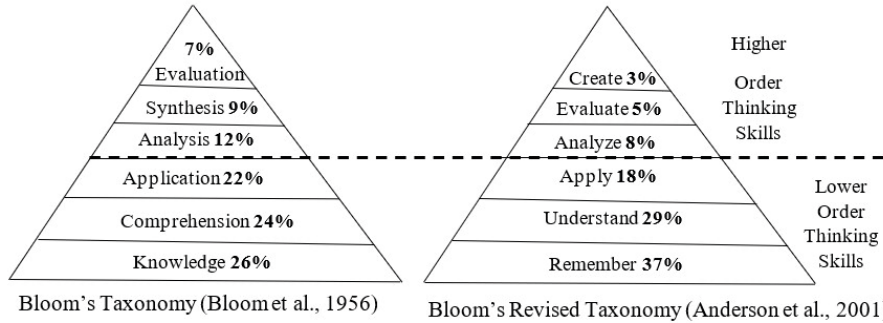
Their complete taxonomy has three major parts such as cognitive, affective, and psychomotor domains. The cognitive domain, which constitutes the six thinking levels, is most emphasised as critical thinking skills and most referred to in curriculum development to establish educational goals. Bloom's Taxonomy serves as a framework for educational objectives and as a standard assessment checklist for evaluating learner learning outcomes. Their six major categories of the cognitive domain start with an essential requirement in learning, *knowledge*. After acquiring the *knowledge*, learners are expected to *comprehend* the subject under study. The next improved level after *knowledge* and *comprehension* is *application* of the knowledge learned. These three basic levels are classified as lower order thinking skills. *Analysing*, *synthesising*, and *evaluating* the existing body of knowledge are three higher order thinking skills that learners are expected to attain as an ultimate learning goal.

In 2001, forty-five years after Bloom's Taxonomy, Anderson and colleagues modified the original taxonomy into a new thinking taxonomy called the Revised Taxonomy or Bloom's Revised Taxonomy. The primary distinction between the revised taxonomy and the original taxonomy is that the noun *Knowledge* is replaced by the verb *Remember*. Krathwohl (2002) explained why the knowledge of any subject matter the learners pursue should be targeted by the action verb that in fact will be the goal to achieve. For example, for an economics student, if the required content knowledge is the law of supply and demand, then the goal is to be able to remember that knowledge. In other words, the reason for modifying the *Knowledge* into *Remember* is to clarify that learners' acquired skill is not the knowledge itself, but the ability or effort to remember or to recall that knowledge.

There are two significant changes between the two taxonomies. The first change in revised taxonomy from the original taxonomy is converting the classification levels of nouns into verb categories. The second change is that

Synthesis skill from the original taxonomy is switched with evaluation level and renamed *Create* in the revised taxonomy. It is not surprising to find out in this review that the two taxonomies are the most frequently used frameworks. It is because of their prominent contribution to the classification of curricular objectives and learning programs since their development hitherto. More importantly, researchers' belief and application of theoretical frameworks is tied to their specialised disciplinary orientation as noted by Merriam and Tisdell (2016). The two taxonomies—which were originally developed for assessing curriculum developments and educational outcomes—become the lenses that textbook evaluation analysts used as critical thinking frameworks to evaluate the potential of the textbooks. However, none of the 15 studies using Bloom's Taxonomy or the 20 studies using the revised taxonomy mentioned any detailed explanation for their choice of one taxonomy over the other, or why the taxonomy they applied was particularly suited to their study compared to their counterpart. Such a lack of clarification on the appropriate choice of the taxonomic framework should be paid attention to in future studies. A critical comparison of the two taxonomies with an evidence-based discussion of teaching materials evaluation will be beneficial.

Targeted Language Levels and Content of the Evaluated Textbooks

Knowing the target language levels of the textbooks allows future researchers to pay more attention to less emphasised areas and compare results across various target language levels. Table 2 shows the number of studies and the language levels of the textbooks being examined.

Table 2

Levels of Target Language in Textbooks Analysed

| Levels of the Target Language	Junior High School & High School	University	Comparison of High School & University	Not Given	Others (Multi-level)	Total
Number of Studies	22	7	2	4	6	41
Percentages	53%	17%	5%	10%	15%	100%

As shown in Table 2, junior high school and high school levels receive the most attention among the studies. There are relatively fewer studies at the university level than at the high school level. Two textbook evaluation studies (Freahat & Smadi, 2014; Riazi & Mosalanejad, 2010) attempted to compare the rate of integration of critical thinking at the high school and college level

for the purpose of seeking transfer patterns and consistency of critical think-ing activities embedded in the textbooks. Such studies contribute to the design and evaluation of curricula at different language levels. Four textbook evalu-ation studies (Birjandi & Alizadeh, 2013; Razmjoo & Kazempourfard, 2012; Sahragard & Alavi, 2016; Ulum, 2016) did not mention sufficient information for the level of textbooks. More complete information on the use of materials and methods would have been helpful for researchers of systematic literature review studies to calculate and integrate the actual existing literature.

Figure 3

Units of Analysis in the textbook Evaluation Studies
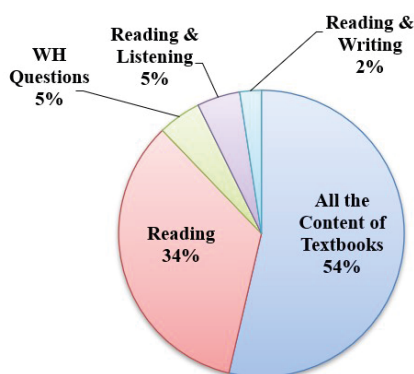
In our analysis, in addition to the level of textbooks, we examined the lan-guage domains of the textbooks being evaluated. Figure 3 shows the language domains of the textbooks analysed for critical thinking integration. There are a total of 125 textbooks distributed among 41 studies, and the different text-books have different language foci and orientations. Therefore, the percentage of the language area for evaluation is influenced not only by the theoretical standpoint and choice of the researchers but simply by the fact that different textbooks have different language areas being focused on such as reading, writ-ing, vocabulary, etc. To this end, it is safe to conclude by looking at Figure 3 that the majority of the studies (54%) indiscriminately analysed the entire content of the textbooks under study. Apart from that, reading received more attention (34%) for critical thinking integration analysis than other language skills. Despite a small percentage, it is noteworthy that two studies (Freahat & Smadi, 2014; Igbaria, 2013) focused on "WH-questions" of the given texts with the theoretical standpoint that questions motivate learners to solve problems and activate learners' critical thinking skills. Asking questions plays a central role in the reinforcement of critical thinking in language learning exercises. Questions asked in a teaching context can be classified as lower level ques-

tions where students need to gather and recall the information, and higher level questions where students need to apply such activities as analysing, synthesising, and evaluating (Nappi, 2017). Teachers' effective use of questions in language exercises can direct students to more strategic and critical thinking.

Validity and Reliability of the Studies

Methodologically valid and reliable research or *trustworthiness* of a study is a broad concept, referring to such factors as *credibility*, *transferability*, *dependability*, and *confirmability* (Shenton, 2004). It ultimately represents how a particular study is conducted methodologically and ethically, so that readers of the study can confidently apply its findings and recommendations to real world problems and can establish social policy in relevant contexts. In this study, we put our focus on two aspects of the coding processes, *validity* and *reliability*, which are critical aspects in the qualitative content analysis research design that the textbook analysts applied. At this point, we refer to Potter and Levine-Donnerstein's (1999) concept of *validity* and *reliability* in content analysis. They underscored the two-step process in content analysis: developing a coding scheme congruent with the theoretical framework of the study, and assessing coders' correct decision making according to certain guidelines and standards. Transparent and elaborative descriptions of the coding process of qualitative content analyses can show its methodological integrity and increase its replicability. This essential characteristic was not presented as favoured in all the reviewed studies. Most researchers did not provide transparent and thorough information about the development of the coding scheme and did not assess how the co-coders make decisions when categorizing critical thinking skills from the analysed textbook content.

In terms of validity, when we examined how many of the studies mentioned the validity in the analysis processes, we found only seven studies (17%) did so. Those seven studies reported that researchers consulted with the experts on their established operational definitions for each critical thinking skill standard and on the appropriateness of the choice of their measurement instrument or critical thinking framework used in their textbook analyses. For example, in, Assaly and Smadi (2015) and in Igbaria (2013), the researchers constructed a list of critical thinking skills containing definitions and concepts of six critical thinking levels from Bloom's taxonomy and presented it to a committee of experts from their specialised field prior to analysis to validate the operational definition of the research instrument. Similarly, Abdelrahman (2014), by consulting a jury of seven experts, validated the suitability of using Bloom's revised taxonomy as a research indicator in assessing critical thinking integration in the analysed textbooks. It is important to have a mutually understood

operational definitions among evaluators on evaluative criteria of the critical thinking skills used to be assessed. This is underscored by Tomlinson and Masuhara (2004) who proposed a list of five questions in a materials evaluation checklist. One of these five questions was to check whether the evaluation criteria were reliable in the sense that other evaluators would define and interpret them in the same way. Although the authors of those studies noted the validity and appropriateness of their choice of research instruments, none of them elaborated on the process of consultation with experts or provided transparent explanations or arguments as to why they chose one particular critical thinking framework over other frameworks.

Figure 4

Instances of Reliability Assessment
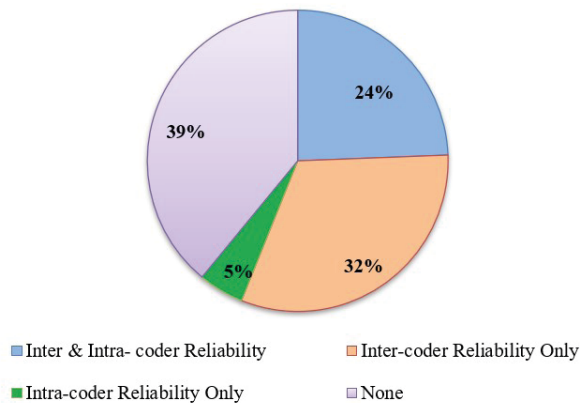
Assessing the reliability of the content analysis is to look at the accuracy and consistency of the coding analysis in the analysis process. In other words, it means finding similar coding judgements about the same content between or among the coders. Figure 4 shows the number of studies that assessed the degree of consistency in their analyses. In reliability assessment, there are inter-rater reliability and intra-rater reliability. By inter-rater reliability, the same content is independently coded by the two or more coders and the findings are compared for agreement (Armstrong et al., 1997). One's own consistent and reliable coding practices are as important as consistent reliability among multiple coders. Therefore, sometime after the initial analysis, the same coder re-analyses the same content. This is called intra-coder reliability assessment. As shown in Figure 4, the majority of the studies, to be precise, 16 studies (39%) did not undergo any reliability assessment in the coding process. Ten studies (24%) applied both inter-coder and intra-coder reliability. Thirteen studies (32%) found the reliability agreement between the two coders. Two of the

studies (5%) sought only intra-coder agreement. It is also found that some of the studies that conducted the reliability assessment did not attempt to assess the full content of the analysis. For example, four studies in the intra-coder reliability assessment group acknowledged that only a small portion of the whole text was randomly selected to check the reliability agreement. In the inter-coder reliability assessment group, five studies acknowledged that only a few portions of the whole text were randomly selected and analysed by their co-coders.

These findings indicate that there is not enough interest or awareness among researchers in the reviewed studies regarding the validity and reliability of assessment methods. There is also little description of the detailed procedures of data handling and analysis methods, which would have been useful to mention for readers seeking to replicate the studies. For such an inadequate methodological description, as Elo and Kyngäs (2008) noted in their methodological guidelines for content analysis research method, researchers need to provide readers with a clear description of how the analysis is done. Overall, these methodological weaknesses reduce the potential value of the results and the reproducibility of the studies. Future studies should be improved in these major aspects of content analysis, validity, and reliability. For the deeper understanding of content analysis research method and its validity establishment and reliability assessment, textbook analysts are referred to Krippendorff (2019).

Recommendations Made by the Studies

Of the three types of textbook evaluations presented in the literature review section of this study, pre-use, in-use, and post-use evaluations, all 41 empirical studies belong to the in-use evaluation phase because the textbooks analysed were already in use at the time of the analysis. None of the studies were in the post-use evaluation phase, as they did not evaluate the learning outcomes after the textbooks were used. Being in-use evaluation studies, their main recommendations were aimed at language teachers who are the primary users of language textbooks. The recommendations are primarily made for effective use of critical thinking enhanced activities in the class and incorporating higher order thinking skills in the textbooks. The following is a summary of the most common recommendations voiced by the textbook evaluation analysts.

- There should be a balance integration of both higher and lower order thinking skills.
- Teachers should add supplementary activities to reinforce learners' critical thinking activities.
- Teachers should be able to evaluate the textbooks they use in language classes.
- Course book writers and teachers should be aware of the importance of critical thinking activities in language teaching materials.

- There should be workshops and training courses to raise this awareness.
- More future studies should be done to investigate other textbooks used in language teaching programs.
- Textbooks should be analysed across the levels of the same series to see the relevancy and consistency of the thinking activities included.
- Teachers' guide books and students' workbooks are overlooked, which also should be evaluated.
- Decision makers in language programs should carefully select language textbooks that fully incorporate higher-order thinking activities.

To summarise the recommendations given, the goals are to include higher-order thinking skills in more language textbooks, to train teachers with the pedagogical knowledge to incorporate critical thinking into textbooks, and to enable teachers to evaluate textbooks from critical thinking perspective. Although these recommendations are reasonable, evidence-based recommendations as the result of post-use textbook evaluation studies will reveal actual classroom experiences more effectively than textual content analyses that do not involve student-teacher interaction and the assessments of learners' performances. Such kind of post-use evaluations tend to be labour-intensive and time-consuming because they are longitudinal studies. However, it is not impossible for language teachers and researchers to apply such post-use evaluation method. Ellis (1997) recommended seven feasible steps for language teachers to evaluate language teaching materials. Ellis's post-use evaluation methods for evaluating language textbooks indicated that language teachers play an autonomous role in this responsibility. Ultimately, it is essential that teachers know the potential contribution of the language materials they are using. However, we should note that since textbooks used in each language program are different, the practical application of textbook evaluation research is limited to specific context. Therefore, it is important to analyse the language textbooks applied in each context.

Limitations of the Study

The three main limitations of this systematic literature review study are the peer review status of the reviewed studies, the limited scope of inclusion criteria in the literature search, and the lack of a full discussion of all critical thinking frameworks applied in the study. In order to incorporate many studies, the inclusion criteria for the literature research was not limited to peer-reviewed papers. However, the number of non-peer-reviewed papers in our analysis was very small compared to that of the peer-reviewed papers. Furthermore, we did not exclude non-peer-reviewed studies because we believe that lessons can be learned from the limitations and weaknesses of existing studies to improve

future research. During the literature review process, we found out that authors in the same EFL context tend to refer to each other's work in their studies. Thus, by pointing out the current status, strengths, and limitations of existing research in the EFL context regardless of being peer-reviewed or not, future researchers can recognise the contributions, limitations, and shortcomings of previous studies. For this reason, we did not exclude non-peer-reviewed papers. Of the critical thinking frameworks, we focused on the two most frequently applied ones by providing background information on these two frameworks. The purpose of this literature review was to examine frequently applied critical thinking frameworks, so we did not provide a thorough description of the remaining applied thinking frameworks. For detailed information, we refer the concerned reader to the work of educational psychologist Moseley et al. (2005), which is a systematic literature review on critical thinking frameworks developed by scholars in various fields over the 50 years prior to 2005. In their thorough review, they found that 41 thinking taxonomies are useful for developing critical thinking in students. These 41 thinking frameworks were then comprehensively described and evaluated by classifying them into three categories: (1) thinking frameworks related to instructional design, (2) thinking frameworks related to productive thinking, and (3) thinking frameworks related to cognitive structure and development. Of these, there are also 13 thinking frameworks related to instructional design for the purpose of curriculum planning and learning assessment. While a review of these critical thinking frameworks is beyond the scope of this study, we suggest that researchers in related fields consider the application of these critical thinking frameworks and conduct analyses comparing and contrasting them.

Conclusion and Recommendations

The present systematic literature review study on ELT textbook evaluation within the period of 12 years delivers a comprehensive overview of the existing literature on the integration of critical thinking enhanced activities and tasks in English language teaching materials from a textbook analysis perspective. Findings of this review study can be scrutinised in relation to the five major units of analysis specified in the review protocol: (1) critical thinking frameworks applied in the studies, (2) the integration percentage of critical thinking, (3) levels and language areas focused on in the studies, (4) validity and reliability of the studies, and (5) recommendations made by the studies. These five major findings can be summarised as follows.

1. For the application of critical thinking frameworks, Bloom's Taxonomy and Bloom's Revised Taxonomy are the most commonly applied frameworks among ELT textbook evaluation studies.
2. As for the integration percentage of critical thinking, lower order thinking skills are overemphasised in the textbooks. So, language activities that foster higher-order thinking skills should be more incorporated into textbooks.
3. For the language levels and language areas being analysed, textbook evaluations are more frequent at junior high school and high school levels than at elementary and college levels.
4. In many studies, methodological validity and reliability assessment of textbook content analysis have not been carefully considered.
5. Reviewed studies mainly recommended that higher and lower order thinking skills should be integrated into balance inside the textbooks.

By reporting on the above findings, this study can benefit two groups of readers. On the one hand, this study can benefit pedagogues, curriculum developers, and stakeholders who are in a position to implement evidence-based policymaking in English language education. Another audience that would benefit from this study is individual researchers and teachers attempting to evaluate ELT materials. Based on the presented literature background, future researchers can establish a research design that is methodologically reliable and valid, especially to compensate for the weaknesses of previous studies.

Future studies should pay attention to the ethical reporting standards (for more detailed ethical reporting standard, see Publication Manual of the American Psychological Association, 2020). From the planning stage of the research project, judgement calls for inclusion and exclusion of what to analyse should be transparently developed so that the replicability of the studies can be enhanced. Data coding and analysis processes of textbook evaluations should be clearly reported because this information is important to the reader to determine the credibility of reported findings and research methods applied. As Nicholls (2003) emphasised in his "Methods of School Textbook Research," the basic elements that textbook researchers should focus on are the methods used to analyse textbooks and the used in the measurement criteria analysis process. We also recommend that an insightful analytical comparative study between the two educational taxonomies is needed so that textbook analysts can determine which theoretical framework is more appropriate for their research purposes. Evaluation studies of ELT textbooks at the college level should receive more attention since they are less prevalent than at the high school level. Examining a series of textbooks prescribed across different target language levels could reveal the pattern and sequencing of critical thinking activities, and could observe if the tasks and activities are developmentally appropriate to learners' cognitive levels. Content analyses of the reading texts should also focus on the topics and inner meaning of the texts because the degree and cognitive complex-

ity of critical thinking activities largely depend on the type of the text. Some texts require learners to possess only basic comprehension skills, while others require learners to possess deep critical thinking. So, the nature and the topic of the content should be included as one factor in ELT textbook evaluation. For that reason, more in-depth multi-layered analysis studies are required. For the last point, we report that two duplicate publications were detected when collecting studies for this systematic review. Such intentional self-plagiarism can jeopardise data computation and interpretation in systematic review studies. We suggest that future researchers avoid such self-plagiarism and pay attention to ethical reporting standards.

References

References marked with an asterisk indicate studies included in the systematic literature review.

*Abdelrahman, M. (2014). An analysis of the tenth grade English language textbooks questions in Jordan based on the revised edition of Bloom's taxonomy. *Journal of Education and Practice, 5*(18), 139–151.

*Adli, N., & Mahmoudi, A. (2017). Reading comprehension questions in EFL textbooks and learners' levels. *Theory and Practice in Language Studies, 7*(7), 590–595. https://dx.doi.org/10.17507/tpls.07074

*Akrong, B. S., Chaka, C., & van der Walt, C. (2021). A modified version of Bloom's taxonomy and critical language awareness: Evaluating EAL textbooks for junior high schools in Ghana through a critical literacy approach. *International Journal of Applied Linguistics and English Literature, 10*(4), 106–115.

*Alfadda, H., Fatima, M., Ghaffar, A., & Afzaal, M. (2020). Critical thinking perspective in ESL new English textbooks: A case study of Pakistan. *International Journal of Applied Linguistics and English Literature, 9*(2), 24–31.

*Alfaki, I. (2014). Sudan English language syllabus: Evaluating reading comprehension questions using Bloom's taxonomy. *International Journal of English Language Teaching, 2*(3), 53–74.

American Psychological Association. (2020). *Publication manual of the American psychological association* (7th ed.). American Psychological Association.

Alnofaie, H. (2013). A framework for implementing critical thinking as a language pedagogy in EFL preparatory programmes. *Thinking Skills and Creativity, 10*, 154–158. http://dx.doi.org/10.1016/j.tsc.2013.09.002

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E, Pintrich, P., Raths, J., & Wittrock, M. C. (2001). *Taxonomy of learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (2nd ed.). Addison and Wesley.

Armstrong, D., Gosling, A., Weinman, J., & Marteau, T. (1997). The place of inter-rater reliability in qualitative research: An empirical study. *Sociology, 31*(3), 597–606.

*Askaripour, S. A. A. (2014). A textbook evaluation of new version of *Top Notch* English series. *English for Specific Purposes World, 15*(44), 1–30.

*Assaly, I. R., & Smadi, O. M. (2015). Using Bloom's taxonomy to evaluate the cognitive levels of master class textbook's questions. *English Language Teaching, 8*(5), 100–110.

*Azizi, M. (2012). Critical thinking in General English course books taught in Iranian Universities. *Journal of Language, Culture, and Translation, 1*(3), 15–36.

Behar-Horenstein, L. S., & Niu, L. (2011). Teaching critical thinking skills in higher education: A review of the literature. *Journal of College Teaching & Learning, 8*(2), 25–42.

*Birjandi, P., & Alizadeh, I. (2013). Manifestation of critical thinking skills in the English textbooks employed by language institutes in Iran. *International Journal of Research Studies in Language Learning, 2*(1), 27–38.

Bloom, B., Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.* Longman.

Brown, H. D., & Lee, H. (2015). *Teaching by principles.* Pearson Education.

Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society, 1*(1), 104–126.

Cottrell, S. (2017). *Critical thinking skills: Effective analysis, argument and reflection.* Bloomsbury Publishing.

Cummins, J. (1999). BICS and CALP: Clarifying the distinction. Washington, DC: Opinion Papers. (ERIC Document Reproduction Service No. ED 438 551). Retrieved from https://eric.ed.gov/?id=ED438551

*Dabbagh, A., & Safaei, A. (2019). Comparative textbook evaluation: Representation of learning objectives in locally and internationally published ELT textbooks. *Issues in Language Teaching, 8*(1), 249–277.

Dummett, P., & Hughes, J. (2019). *Critical thinking in ELT: A working model for the classroom.* National Geographic Learning.

Dummett, P. (2016). *How to write critical thinking activities.* ELT Teacher 2 Writer.

*Es-Salhi, A., & Elfatihi, M. (2019). Evaluating critical thinking skills in Moroccan EFL textbooks: *Gateway to English 2* as a case. *Higher Education of Social Science, 17*(1), 13–22. http://doi.org/10.3968/11284

Elder, L., & Paul, R. (2012). Critical thinking: Competency standards essential to the cultivation of intellectual skills, Part 4. *Journal of Developmental Education, 35*(3), 30–31.

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

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing, 62*(1), 107–115.

Ennis, R. H. (2018). Critical thinking across the curriculum: A vision. *Topoi, 37*(1), 165–184. https://doi.org/10.1007/s11245-016-9401-4

Facione, P. A. (2011). Critical thinking: What it is and why it counts. Academic Press. Retrieved from https://courseware.e-education.psu.edu/downloads/geog882/Critical%20Thinking%20What%20it%20is%20and%20why%20it%20counts.pdf

Fahim, M., Barjesteh, H., & Vaseghi, R. (2012). Effects of critical thinking strategy training on male/female EFL learners' reading comprehension. *English Language Teaching, 5*(1), 140–145.

*Fitriani, D. H., & Kirana, D. P. (2022). Higher order thinking skill (hots) in English language textbook in senior high school. *International Journal of Research on English Teaching and Applied Linguistics, 2*(2), 55–66.

*Freahat, N. M., & Smadi, O. M. (2014). Lower-order and higher-order reading questions in secondary and university level EFL textbooks in Jordan. *Theory & Practice in Language Studies, 4*(9), 1804–1813. https://doi.org/10.4304/tpls.4.9.1804-1813

Gardner, H. (1993). *Frames of mind: The theory of multiple intelligences.* Basic Books.

Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. Basic Books.

Golpour, F. (2014). Critical Thinking and EFL learners' performance on different writing modes. *Journal of Pan-Pacific Association of Applied Linguistics, 18*(1), 103–119.

*Gordani, Y. (2010). An analysis of English textbooks used at Iranian guidance schools in terms of Bloom's taxonomy. *Journal of Asia TEFL, 7*(2), 249–278.

Halpern, D. F. (1999). Teaching for critical thinking: Helping college students develop the skills and dispositions of a critical thinker. *New Directions for Teaching and Learning*, 89, 69–74.

Halpern, D. F. (2002). Teaching for critical thinking: Helping college students develop the skills and dispositions of a critical thinker. *New Directions for Teaching and Learning, 1999*(80), 69–74.

*Hei, J. X. (2017). Thinking skills in the 12th grade English course book in Timor Leste. *Beyond Words, 5*(1), 32–45.

Hughes, J. (2014). Critical thinking in the language classroom. *Recanati (Italy): ELI*. Retrieved from: http://www.elionline.com/eng/teachers-area/critical-thinking

Ilyas, H. P. (2015). *Critical thinking: Its representation in Indonesian ELT textbooks and education* [Unpublished doctoral thesis]. University of York.

*Igbaria, A. K. (2013). A content analysis of the WH-Questions in the EFL textbook of *Horizons. International Education Studies, 6*(7), 200–224. https://doi.org/10.5539/ies.v6n7p200

*Jebbour, M. (2019). Exploring the manifestation of critical thinking in the Moroccan textbooks of English: The case of *Ticket 2 English. JEES (Journal of English Educators Society), 4*(1), 39–45.

Jones, E., Hoffman, S., Moore, L., Ratcliff, G., Tibbetts, S. & Glick, B. (1995). *Essential Skills in Writing. Speech and Listening. and Critical Thinking for College Graduates: Perspectives of Faculty, Employers, and Policy Makers*. Project summary, U.S. Department of Education, OERI Contract No. R117GI0037. National Center for Post-secondary Teaching, Learning and Assessment. University Park, Pennsylvania State University Press.

*Khodadady, E., & Karami, M. (2017). An evaluation of textbooks designed for advanced English learners within a foreign context: A critical thinking perspective. *Porta LINGUARUM, 28*, 95–109.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 41*(4), 212–218.

Krippendorff, K. (2019). *Content analysis: An introduction to its methodology*. Sage.

Lin, Y. (2018). *Developing critical thinking in EFL classes*. Springer.

Littlejohn, A. (2011). The analysis of language teaching materials: Inside the Trojan Horse. In B. Tomlinson (Ed.), *Materials development in language teaching* (2nd ed., pp. 179–211). Cambridge University Press.

Lune, H., & Berg, B. L. (2017). *Qualitative research methods for the social sciences*. Pearson.

*Margana, & Widyantoro, A. (2017). Developing English textbooks oriented to higher order thinking skills for students of vocational high schools in Yogyakarta. *Journal of Language Teaching and Research, 8*(1), 26–38. http://dx.doi.org/10.17507/jltr.0801.04

McDonough, J., Shaw, C. & Masuhara, H. (2013). *Materials and Methods in ELT* (3rd ed.). John Wiley & Sons.

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

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation*. Jossey-Bass.

*Mizbani, M., & Chalak, A. (2017). Analyzing reading and writing activities of Iranian EFL textbook *Prospect 3* based on Bloom's revised taxonomy. *Journal of Applied Linguistics and Language Research, 4*(2), 13–27.

Moher D, Liberati A., Tetzlaff J., Altman D., the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine, 151*(4), 264–270.

Moore, T. (2013). Critical thinking: Seven definitions in search of a concept. *Studies in Higher Education, 38*(4), 506–522.

Moore, B. & Parker, R. (2009). *Critical thinking* (9th ed.). McGraw-Hill Higher Education.

*Muchlis, M. (2015). An analysis of thinking order of reading comprehension questions in English textbook for young foresters of Forestry Vocational School of Samarinda. *Jurnal Nalar Pendidikan, 3*(1), 314–316.

*Nainggolan, B. R., & Wirza, Y. (2021). Evaluation of an English textbook: Does *English in Mind* promote critical thinking skills? *Advances in Social Science, Education and Humanities Research, 546*, 51–59.

Nappi, J. S. (2017). The importance of questioning in developing critical thinking skills. *Delta Kappa Gamma Bulletin, 84*(1), 30–41.

Nicholls, J. (2003). Methods in school textbook research. *International Journal of Historical Learning, Teaching and Research, 3*(2), 1–17.

Nilson, L. B. (2021). *Infusing critical thinking into your course: A concrete, practical approach.* Stylus Publishing.

*Olimat, M. (2015). Analyzing action pack textbooks' questions according to revised Bloom taxonomy. *Journal of Education and Practice, 6*(28), 152–159.

Peterson, S. (2008). *Master critical thinking for the SAT.* Peterson's.

Potter, W. J., & Levine-Donnerstein, D. (1999). Rethinking validity and reliability in content analysis. *Journal of Applied Communication Research, 27*, 258–284.

Puig, B., Blanco-Anaya, P., Bargiela, I. M., & Crujeiras-Pérez, B. (2019). A systematic review on critical thinking intervention studies in higher education across professional fields. *Studies in Higher Education, 44*(5), 860–869.

*Rahpeyma, A., & Khoshnood, A. (2015). The analysis of learning objectives in Iranian junior high school English textbooks based on Bloom's revised taxonomy. *International Journal of Education & Literacy Studies, 3*(2), 44–55. https://doi.org/10.7575/aiac.ijels.v.3n.2p.44

*Raqqad, Y. M. A., & Ismail, H. H. (2018). Analyzing the reading questions of AP 12 textbook according to analyzing the reading questions of *ap12* textbook according to Bloom's taxonomy. *International Journal of Education*, *Psychology and Counselling, 3*(22), 84–94.

*Razmjoo, S. A., & Kazempourfard, E. (2012). On the representation of Bloom's revised taxonomy in *Interchange* course books. *The Journal of Teaching Language Skills (JTLS), 4*(1), 171–204.

*Riazi, A. M., & Mosalanejad, N. (2010). Evaluation of learning objectives in Iranian high-school and pre-university English textbooks using Bloom's taxonomy. *TESL-EJ, 13*(4)1–16.

*Roohani, A., Taheri, F., & Poorzangeneh, M. (2013). Evaluating *Four Corners* textbooks in terms of cognitive processes using Bloom's revised taxonomy. *Research in Applied Linguistics, 4*(2), 51–67.

*Roohani, A. (2015). Analyzing cognitive processes and Multiple Intelligences in the nop-notch textbooks. *Journal of Modern Research in English Language Studies, 2*(3), 65–39.

*Sahragard, R., & Zahed Alavi, S. (2016). Investigating the predominant levels of learning objectives in general English books. *Journal of English Language Teaching and Learning, 17*(8), 93–114.

Sanavi, V. R., & Tarighat, S. (2014). Critical thinking and speaking proficiency: A mixed-method study. *Theory and Practice in Language Studies, 4*(1), 79–87.

Sheldon, L. E. (1988). Evaluating ELT textbooks and materials. *ELT Journal, 42*(4), 237–246.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information, 22*(2), 63–75.

Shirkhani, S., & Fahim, M. (2011). Enhancing critical thinking in foreign language learners. *Procedia-Social and Behavioral Sciences, 29*, 111–115. https://doi.org/10.1016/j.sbspro.2011.11.214

*Shuyi, N. S., & Renandya, W. A. (2019). An analysis of the cognitive rigour of questions used in secondary school English language textbooks in Singapore. *Asian Journal of English Language Studies (AJELS), 7*, 169–188.

*Sirait, R. E. S. R. (2014). Comprehension levels of reading exercises in *Look Ahead English* coursebooks. *International Journal of English Language Education, 2*(2), 253–268. http://dx.doi.org/10.5296/ijele.v2i2.6307

Solihati, N., & Hikmat, A. (2018). Critical thinking tasks manifested in Indonesian language textbooks for senior secondary students. *Sage Open, 8*(3), 1–8. https://doi.org/10.1177/2158244018802164

*Sucipto, S., & Cahyo, S. D. (2019). A content analysis of the reading activities in *Bright 2* an English textbook for junior high school students. *English Language Teaching Educational Journal, 2*(1), 13–21.

*Tabari, M. A., & Tabari, I. A. (2015). Links between Bloom's taxonomy and Gardener's Multiple Intelligences: The issue of textbook analysis. *Advances in Language and Literary Studies*, 6(1), 94–101. https://doi.org/10.7575/aiac.alls.v.6n.1p.94

*Talebinezhad, M. R., & Matou, Z. (2012). EFL reading comprehension textbooks at university level: A critical thinking perspective. *Theory & Practice in Language Studies, 2*(8), 1746–1754. https://doi.org/doi:10.4304/tpls.2.8.1746-1754

*Tangsakul, P., Kijpoonphol, W., Linh, N. D., & Kimura, L. N. (2017). Using Bloom's revised taxonomy to analyze reading comprehension questions in *Team Up* in English 1–3 and grade 9 English o-net tests. *International Journal of Research-Grantaalayah, 5*(7), 31–41.

*Tayyeh, M. N. (2021). An analysis of reading comprehension questions in English textbook *English For Iraq* according to revised Bloom's taxonomy. *Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12*(7), 2868–2874.

Thomson, C. (2011). Critical thinking across the curriculum: Process over output. *International Journal of Humanities and social science, 1*(9), 1–7.

Tomlinson, B. (2012). Materials development for language learning and teaching. *Language Teaching, 45*(2), 143–179.

Tomlinson, B., & H. Masuhara (2004). *Developing language course materials*. RELC Portfolio Series.

*Ulum, Ö. G. (2016). A descriptive content analysis of the extent of Bloom's taxonomy in the reading comprehension questions of the course book Q: Skills for success 4 reading and writing. *The Qualitative Report, 21*(9), 1674–1683.

*Wu, Y., & Pei, Z. (2018). An investigation of critical thinking manifested in the questions of EFL textbooks for tertiary-level English majors of China. *American Journal of Education and Learning, 3*(2), 72–84. https://doi.org/10.20448/804.3.2.72.84

Webb, N. L. (2002). *Depth of knowledge levels for four content areas* [Unpublished manuscript]. Wisconsin Center for Education Research, University of Wisconsin Madison. Retrieved from http://ossucurr.pbworks.com/w/file/fetch/49691156/Norm%20web%20dok%20by%20subject%20area.pdf

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

Xiao, Y., & Watson, M. (2017). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research, 39*(1), 93–112. https://doi.org/10.1177/0739456X17723971

Yang, Y. T. C., & Gamble, J. (2013). Effective and practical critical thinking-enhanced EFL instruction. *ELT Journal, 67*(4), 398–412.

*Zamani, G., & Rezvani, R. (2015). 'HOTS' in Iran's Official Textbooks: Implications for Material Design and Student Learning. *Journal of Applied Linguistics and Language Research, 2*(5), 138–151.

*Zareian, G., Davoudi, M., Heshmatifar, Z., & Rahimi, J. (2015). An evaluation of questions in two ESP course books based on Bloom's new taxonomy of cognitive learning domain. *International Journal of Education and Research, 3*(8), 313–326.

Thiri Soe

Untersuchung des kritischen Denkens in Englischlehrbüchern: Eine systematische Literaturübersicht über Studien zur Evaluation von Lehrbüchern

Zusammenfassung

Die zentrale Rolle des kritischen Denkens und dessen Integration in den Englischunterricht sind durch systematische Literaturübersichten und Metaanalysen in den Bereichen von pädagogischen Interventionen, Lehrmethoden und Bewertungsmethoden zusammengefasst. Es gibt allerdings kaum Belege für eine systematische Literaturübersicht über die Integration von kritischem Denken in Lehrmaterialien für den Englischunterricht, wie z. B. Lehrbücher. In der vorliegenden Arbeit wurden 41 empirische Studien zur Evaluation von Lehrbüchern aus den Jahren 2010-2021 im Hinblick auf kritisches Denken untersucht. Ziel der Arbeit ist es, einen Überblick über die Evaluation von Englischlehrbüchern in Bezug auf die Integration von kritischem Denken zu geben, die angewandten Rahmenkonzepte für kritisches Denken zu identifizieren und mögliche Forschungslücken aufzuzeigen. Hierfür wurde eine umfangreiche Recherche unter Anwendung der achtstufigen systematischen Literaturübersichtsmethode von Xiao und Watson (2017) durchgeführt. Basierend auf einer Sammlung von 41 empirischen Studien konzentriert sich die Arbeit auf folgende vier Aspekte: angewandte Rahmenkonzepte für kritisches Denken, Untersuchungsergebnisse, Empfehlungen sowie Sprachniveaus und Inhaltsbereiche der untersuchten Lehrbücher. Die Ergebnisse zeigen, dass (1) die Bloomsche Taxonomie und die revidierte Bloomsche Taxonomie die am häufigsten angewandten Rahmenkonzepte für kritisches Denken waren, (2) die Integration von Denkfähigkeiten höherer Ordnung in Lehrbüchern seltener war als die von Denkfähigkeiten niedrigerer Ordnung, (3) die Evaluation von Lehrbüchern häufiger in der Mittel- und Oberstufe erfolgte als in anderen Stufen und (4) methodische Überlegungen zur Reliabilität und Validität des Kodierungsprozesses bei der Inhaltsanalyse von Lehrbüchern wenig Beachtung fanden. Die Arbeit liefert einen zusammenfassenden Literaturüberblick über die Evaluation von Englischlehrbüchern mit Empfehlungen für methodische Verbesserungen in zukünftigen Studien.

Schlüsselwörter: systematische Literaturübersicht, kritisches Denken, Evaluation von Englischlehrbüchern, Denkfähigkeiten höherer Ordnung, Denkfähigkeiten niedrigerer Ordnung