ACADEMIA International Journal for Social Sciences
Volume 4, Issue 3, 2025 ISSN-L (Online): 3006-6638

A Comprehensive Review of Reading Comprehension Assessment Methods and Cognitive Levels

Ishfaque Ali Kalhoro
ishfaquealikalhoro@gmail.com
Lecturer in English at the Shaikh Ayaz University, Shikarpur, Pakistan

Fazal Rabbi
fzrabbi62@gmail.com
English Department, Govt College of Nursing Saidu Sharif Swat

Bait Ul Haram
mehakfaisal83@gmail.com
Nursing Instructor, Govt College of Nursing Saidusharif, Swat
Corresponding Author: * Ishfaque Ali Kalhoro ishfaquealikalhoro@gmail.com

Received: 12-05-2025 **Revised:** 25-06-2025 **Accepted:** 10-07-2025 **Published:** 28-07-2025

ABSTRACT

This review article aims to analyze the essential aspects of measuring reading proficiency. The assessment considers crucial variables, including the selection of text for evaluating reading comprehension, the structure of items, the length of the text, the inclusion of different genres, and the importance of previous knowledge in assessing students at different levels or stages. Most of the studies employed several taxonomies to evaluate reading comprehension. Many researchers have widely used Bloom's taxonomy to evaluate lower and higher-level skills. The data was collected from many sources, including books and journals. The previous study's findings indicated that the design of reading tests was significantly dependent on the objective of evaluating specific abilities or subskills at students' higher and lower cognitive levels. Various aspects were considered, including the test-takers' skill levels, the texts' choice and length, the genre of texts, the forms of the test items, and the applicants' previous knowledge.

Keywords: *Bloom's taxonomy, reading comprehension, testing, test format*

INTRODUCTION

Testing is a method used to assess a test taker's performance, knowledge, or ability (Han, 2022). Brown (2004) defined testing as a method used to assess an individual's competence or comprehension in a certain domain. A method is a set of procedures, techniques, and tools that need an instrument to accomplish an activity the test taker performs. Anaya et al. (2022) state that tests assess students' language skills and encourage them to study. Examinees are assessed on their proficiency in meaningful and purposeful tasks relevant to real-life communication. Testing tools evaluate knowledge and the ability to perform linguistic tasks in real-life scenarios. Assessing comprehension involves considering reading as both a process and an outcome (Catts, 2022). Process refers to the active involvement of the reader with the text. The word "product" refers to the outcomes of the procedure. This seems to be designed to evaluate the proficiency of a test in comprehending a certain text, suitable forms for test items, implementation of a specific research approach, and subsequent examination of the results (Sherkuziyeva et al., 2023). Due to the diverse backgrounds and life experiences of students, it is unavoidable that their reading results would vary.

Reading is considered a crucial academic skill, serving as a method for acquiring new knowledge, especially for those learning English as a second or foreign language. Significant assessments like the

TOEFL, TOEIC, and IELTS are significant to the mastery of reading comprehension abilities. Reading is an interactive activity that simultaneously activates several faculties (Johnson & Lester, 2022). The connection between the reader's prior knowledge and the information presented in the text is a sort of advanced comprehension in this interaction (Hattan, 2024).

Reading comprehension encompasses comprehending explicit information delivered in straightforward language and implicit information that readers infer and connect to prior knowledge (Blaži Ostojić, 2023). While writings include inherent meanings, readers with varying backgrounds, educational levels, and areas of interest sometimes interpret texts differently (Jian, 2022). Furthermore, their interpretations may change as they read the same book in different circumstances. Multiple studies have been conducted to investigate the characteristics of reading skills. Defining reading skill development, training, and assessment is difficult due to contextual variability (Sainsbury et al., 2006).

The assessment of reading comprehension skills is often approached from two contrasting viewpoints: reading as a multifaceted aptitude and reading as a singular capability (Weir & Porter, 1994). Based on previous viewpoints, reading competency does not include separate subskills or subcomponents (Alderson, 1990; Bachman, 1990; Rost, 1993; Weir & Porter, 1994; Shah et al., 2025). Reading combines two interrelated processes: comprehension and word recognition, sometimes called decoding (Perfetti & Helder, 2022). Hughes (2003) examined the challenges of subskills in developing reading abilities. He also emphasized the need to include specific skills in reading comprehension assessment. Brysbaert (2022) argues that evaluating reading skills requires a well-defined reading theory that outlines reading abilities and enables accurate assessment and interpretation of students' performance.

According to the later reading viewpoint (multi-divisibility), a specific reading ability consists of identifiable and unique subskills. Previous studies have effectively shown that a particular reading ability may be divided into several subskills, as evidenced by studies (Davis, 1968; Kim, 2009; Matthews, 990; Munby, 1978; Shah et al., 2022). A comprehensive comprehension of reading abilities and the creation of reliable assessment items and syllabus planning rely on the correlation between high and low-order thinking skills in reading tests and their differentiation (Hughes, 2003; Lumley, 1993; Weir & Porter, 1994). A fundamental challenge that emerges from the multifaceted nature of reading skills is determining the proper sequence of skills or sub-skills when evaluating reading proficiency and the range of abilities. Research on acquiring both first and second languages has not succeeded in hierarchically organizing reading skills. No compelling evidence supports the strict classification of reading skills into high- and low-order tasks (Alderson, 1990; Hudson (2007).

Despite the abundance of data, there is a prevailing inclination towards the importance of reading abilities and subskills. This is apparent in the many taxonomies used in reading teaching and evaluation, as documented by Urquhart (1998), Pearson and Johnson (1978), Alderson and Lukmani (1989), Grabe (1991), Hudson (2007), and Matthews (1990). Although these taxonomies are beneficial, they also have critics. Matthews (1990) argued that Munby's taxonomy (1978) is primarily grounded on knowledge rather than talents. Hudson (2007) argued that Bloom, Barrett, and Munby's taxonomies are mostly based on theoretical assumptions rather than empirical frameworks.

Reading comprehension examinations are impacted by several factors, such as the substance of the text, the questions asked, and the level of knowledge or skill required. The difficulty of reading test items is influenced by several factors, not only the level of competence required (Conrad et al., 2022; Oktarin et al., 2022; Day & Park, 2005; Kobayashi, 2005; Shah et al., 2022). The variables included in this study consist of the item structure, contextual factors, the implicit and explicit characteristics of the content, student strategies during test-taking, the need for higher cognitive abilities, the length of the text, and the kind of text (Mckenna & Stahl, 2015; Pearson & Johnson, 1978). Several factors, including the content and length of the stem, the length of the alternatives, the length of the correct option, and the presence of

distracters, influence the structure of multiple-choice items. Pearson and Johnson (1978) found that the information provided in questions, whether it is stated directly or indirectly, influences the level of difficulty of an item. Hence, to get dependable inferences on high and low-order talents, it is important to consider these factors that influence the difficulty of items. The constructive-integrative process hypothesis, presented by Kintsch (1988), highlights the importance of past knowledge in evaluating reading comprehension.

This study provides a comprehensive examination of the concepts behind the development of reading evaluations. When assessing reading comprehension, two categories of variables are employed: reader variables, which include the diverse levels of knowledge and understanding of the test takers, their background knowledge, and their cultural background, and textual variables, which encompass factors such as the length of the text, the genre, and the regularly used question forms. When used as a practical framework for evaluation (Marzano & Kendall, 2007), Bloom's taxonomy is valuable for assessing learners' abilities since it enables the examination of different kinds of test items based on learners' proficiency levels. This study contributes to the advancement of future tests by incorporating the given principles and assuring the authentic achievement of the testing aim. This review analyzes literature from reputable studies and books to extract the fundamental principles of testing reading comprehension. These principles include guidelines for selecting texts, determining text length, constructing item formats for testing reading at various levels, considering background knowledge, and identifying the skills to be measured in a reading comprehension test.

TEXT SELECTION

The selection of text is crucial when constructing a reading evaluation (Wyse & Bradbury, 2022). Studies suggest two distinct selection criteria: those that relate to students and those that relate to text selection. When defining criteria connected to students, examining their interests, content level, needs, reading aim, and cultural schema (Arias, 2007). Berkoff (1979) suggested that while creating a test, the material should be selected based on the learners' content level, cultural schema, and language level, which includes grammar, vocabulary, and syntactic complexity. Conradi et al. (2022) state that the selection of text influences the reading objective of students. One may evaluate reading skills and subskills using this method. Prior to mandating students to read the assessment material, it is crucial to consider their prior knowledge and provide them with the relevant background information (Ferris & Hedgcock, 2023.

Reading comprehension is evaluated using a variety of assessment questions. Activities include prereading, skimming, scanning, reading comprehension, translating, reading aloud, and engaging with literature. However, reading comprehension is covered in great depth in schools (Gallagher, 2023). He asserted that reading evaluations function as feedback by making readers mentally reproduce a particular text because they force them to respond to it. A number of factors are considered while selecting a text, including its content, appropriateness, readability, subject matter, appearance, and suitability for a particular political and cultural setting. Bishop-Monroe et al. (2022) asserted that students' real-world experiences and professional surroundings should inform their chosen curriculum. The exam material should be chosen with the teacher's objectives for the reading session in mind. It can be exploitable due to the thematic and linguistic content. A book should be suitable, acceptable, and engaging for the test taker (Feng et al., 2023). The concepts in the essay should flow logically, agreeably, and rhetorically.

The discourse level of textual phenomena and the reader's previous knowledge must be acknowledged and honored (Day, 1994). The linguistic qualities of several genres should not cause confusion for the reader (Grabe & Stoller, 2002), and the text should maintain consistency and be easily comprehensible to them (Nuttall, 1996).

GENRE

Genre is a crucial element in the construction of tests. The genre significantly impacts the selection of texts. Different genres are used to different extents. Pham and Bui (2022) emphasized that a reading book should motivate students to use the language in a specific context and provide a practical objective for assessing their language skills. They proposed a range of text genres, including textbooks, newspaper articles, magazines, journals, excerpts from poems, encyclopedic entries, dictionary entries, leaflets, letters, timetables, excerpts from novels, advertisements, diaries, postcards, computer assistance manuals, reviews, maps, and short stories. According to the studies, readers of this kind of literature must use certain schemata to assist them in deducing the correct meaning.

Text style and genre play a key role when evaluating reading comprehension. Texts are categorized into four distinct types: narrative, argumentative, expository, and descriptive. Research suggests that expository writings are challenging to comprehend, but narrative texts are more accessible due to the interconnections between textual elements and their abundant content. The macrostructure of the narratives seems to enhance reader comprehension. Weir (1990) stated that each book is specifically designated for a distinct testing methodology. Ali et al. (2022), have the same perspective, argues that a singular approach may effectively meet the criteria for all assessments or for all written works and abilities that are intended to be evaluated. According to some ELT professionals, the material influences the performance of test-takers. Kuromiya et al. (2022) conducted a study and discovered that Japanese high school students were given two story readings and a typical explanation book as assignments. In addition, he observed a significant disparity in the reading scores between explanatory and narrative literature. It was noted that the reading comprehension abilities of L2 learners were influenced differently by the various kinds of texts, particularly due to the students' strong performance on narrative texts. The organization of texts differs across different genres. Scholars have shown interest in the correlations within paragraphs and the connections between ideas and sequences of events. Researchers examined several organizational frameworks within a certain genre, which had diverse outcomes. Texts may be categorized into four distinct types: explanatory, narrative, descriptive, and argumentative.

Various types of literature need distinct skills from the reader, potentially influencing their degree of understanding. Hughes (2003) categorized different types of texts, including narrative texts, jokes, reports, biographical notes, stories, news, and historical accounts as examples of narrative texts. Expository texts encompass explications, impressionistic descriptions and outlines, summaries, and text interpretations. Argumentative texts consist of instructional texts, personal instructions, comments and formal argumentations, practical instructions, and jokes. Regarding reading assessments, research has shown that other elements influence the difficulty of test items apart from the intrinsic difficulty of the skill itself (Bachman, 1990; Day & Park, 2005; Kobayashi, 2005; McNamara & Roever, 2006). The factors that influence the test include the testing environment, the type and format of the questions, the explicit and implicit information being assessed, the cognitive load, the strategies employed by the students during the test, the type of text being used, and the length of the text.

The format of a multiple-choice question is influenced by several factors, including the length of the question stem, the words used in the stem, the structure of the options, the length of the correct answer, and the distractors. Furthermore, Pearson and Johnson (1978) found that the complexity of an item determines the level of implicitness and explicitness in question information. Brown (2004) highlights and categorizes textual genres since they are the foundation for evaluating reading skills. The writer categorizes written texts into three groups: academic genres, which encompass articles of general interest such as those found in newspapers and magazines. Technical reports like laboratory reports; articles authored by professionals; and reference materials like dictionaries, books, essays, and papers; job-related readings, which consist of financial documents, reports, instructions, and correspondence; and personal

reading, which encompasses books, magazines, jokes, drama, poetry, maps, recipes, greeting cards, invitations, messages, notes, lists, questionnaires, forms, and immigration documents. In order to focus on the text and understand its meaning, the tester must know the governing standards and regulations that each genre creates for the display of the text. People often have a proclivity for thinking in a similar manner. Consequently, readers may make predictions about what to expect from the material. International testing organizations like the IELTS include reading passages from various genres, including narrative, descriptive, and argumentative texts. The TOEFL test incorporates texts from various academic subjects, such as science, art, and psychology.

Similarly, the GRE test utilizes content from novels, textbooks, and biographies. Liu (2011) states that many research studies have shown that expository writings are more challenging to comprehend than narrative texts. This may be attributed to the complex interplay between textual units and the diverse array of content.

TEXT LENGTH

Alderson (2000) argues that test producers have difficulties establishing the appropriate text length for different test levels due to the association between reading comprehension ability and text length. According to him, there is still a lack of research on the relationship between text length and reading comprehension. There is a natural belief that a passage will get more difficult as its length increases. Nevertheless, the many components of the test procedure and the student's level of reading competence significantly influence the student's academic achievement. The test input format consists of many factors, including the text's length, form, channel, and degree of speediness.

On the other hand, the test method components include rubrics, testing setup, and input (Bachman, 1990; Palmer, 1996). When information is conveyed in brief intervals, it must be understood within a more limited scope of potential meanings compared to when it is offered in a lengthier discussion. Alderson (2000) stated that texts may vary in length, ranging from a single word to a phrase (such as warnings and public notices), a single sentence (such as advertising), an entire book, two paragraphs on postcards, or even a twenty-page research study. He also states that the amount of interpretation required is directly proportional to the length of the text. Surprisingly, there is no connection between interpretation and length. For example, although a book may include explicit content, a postcard may provide greater difficulty in terms of comprehension (Alderson, 2000). Heaton (1975) proposed using 50–100 words for the basic level, 200–300 for the elementary level, and 400–600 for the expert level. The length of a text varies based on the genre chosen by the author. A text may consist of excerpts selected from a compilation of occurrences and a text with a length of 2000 words is considered appropriate. According to the Flesch-Kincaid index, paragraphs with 800 to 1200 words are appropriate for readers with advanced skills. Alderson (2000) observed a distinct change in test-takers' measuring abilities when the passage length approached 1,000 words. According to him, lengthy texts assessed students' ability to study and reduced their reliance on understanding sentences inside and between paragraphs.

Consequently, it leads students to prioritize analyzing discourse material above their understanding of sentence structure and vocabulary. Moreover, there might be variations in the ability to identify the fundamental idea in lengthier texts instead of shorter ones. Using longer texts simplifies reading speed measurement compared to shorter texts with associated questions. Andreassen and Braten (2010) found that longer texts may increase the cognitive load on working memory. Extensive texts have also shown the ability to induce mind wandering, negatively impacting the ability to understand and comprehend what is being read. Carver (1990) proposed that the most accurate method for determining the length of a text is to consider that each word requires six characters, including spaces. The longer works had lengths ranging from 600 to 650 words, while the lesser ones ranged from 300 to 350 words. Instead of using

shorter texts, Carter, Rastatter, Walker, and O'Brien (2009) found that using lengthier texts enhanced reading comprehension abilities during testing.

The texts had a length of around one-fourth of a page, and a font size of 12 Times New Roman was used to provide double space on the pages. The results showed that reading longer texts at a quicker pace improved reading comprehension and had a positive additional outcome. However, the current structure for global assessments, such as the GRE (1–5), TOFEL (3–4), and IELTS (3–5), requires the inclusion of three paragraphs. Therefore, considering several criteria, a text length of 3-5 paragraphs might be considered appropriate. Wolfe and Woodwyk (2010) found that reading shorter texts promotes the use of bottom-up methods and requires more concentration when processing words individually. The little information included in concise words makes them more challenging. Conversely, longer paragraphs are preferable since they are more readable and include more significant, overarching information.

Previous study indicates that texts consisting of 3-5 paragraphs are the maximum length suitable for assessing reading comprehension. Conversely, this allows for a broader range of topics to be addressed and enables the control of bias by the test creators. A brief section usually has between 300 and 600 words, whereas longer texts may have a word count ranging from 600 to 1200. Reading evaluations benefit from longer texts since they include a wider range of metacognitive skills, including the ability to understand the main concept, make inferences, and derive meaning from context. They are suitable for readers with high proficiency. A longer phrase benefits the testers by providing them with more context and additional facts. When designing reading assessments, test writers should find a middle ground between reducing the complexity of the content by using several shorter texts and maximizing the authenticity of longer texts that students are already familiar with from their coursework.

TYPES OF TEST FORMATS

Various methods may be used to evaluate reading comprehension, such as the cloze test, multiple choice questions, gap-filling exercises, picture-cued items, scanning, and skimming. Heaton (1990) proposed using matching tests to measure intermediate reading competency. These tests include phrase matching, word matching, picture matching, and sentence matching, which are used to evaluate fundamental reading skills. For higher-level exams, it is recommended to use several types of questions such as multiple choice, true/false, completion, rearrangement, open-ended, and other similar formats. Additionally, the cloze technique and cursory reading may also be used. These approaches are used to assess both low-order and high-order abilities across different skill levels. Brown (2004) proposed many test styles, including written response, reading aloud, multiple choice, matching, picture-cued items, editing, cloze test, gap filling test, C-test, cloze-elide test, ordering test, short-answer exam, and summary test. Alderson (2000) and Cutting (2017) have suggested many approaches for assessment, including multiple-choice, gap-filling, cloze, matching, editing, ordering, cloze-elide, summary, short-answer, gapped summary, and information-transfer.

When choosing an item format, it is important to consider the measurement of content and cognitive processes, as suggested by Haladyna (2004). Hughes (2003) suggests that the task should align with the assessed talents. Kobayashi's (2002) study examined the relationship between test type, test items, and students' test performance. Previous research used three primary examination formats: cloze, summary writing, and open-ended questions. The study's findings indicate a robust association between these criteria and students'' academic performance. According to the study conducted by Zheng, Cheng, and Klinger (2007), different types of tests evaluate different talents of students. No one test style comprehensively examines all abilities in question.

Multiple-choice questions (MCQs) are a very important assessment type for testing reading comprehension. According to Brown (2004), one advantage of implementing this test is its ease of

administration and scoring. The Multiple Choice Question (MCQ) is presently the most versatile and adaptable format for testing. A wide range of learning outcomes, from simple to complex, may be measured. MCQs are very versatile and may be used for a broad range of subjects, so they are only employed in high-stakes exams. A multiple-choice question (MCQ) generally has three components: the exact or precise response, the number of distractors (inappropriate and erroneous replies), and the stem (which presents the problem or issue). Multiple-choice questions (MCQs) are created to test various learning goals, ranging from simple recollection of facts to the most advanced level of cognition defined by Bloom's taxonomy evaluation (Osterlind, 1998). Heaton (1988) argues that multiple-choice testing does not accurately assess language proficiency as a form of communication. However, it is true that correctly and impartially scoring a multiple-choice test is a straightforward task. Contrary to essay writing exams, these evaluations benefit struggling readers and do not impose penalties on them (Chan & Kennedy, 2002).

An inherent problem with multiple-choice questions is that the subjective construction by the test author might lead to a test that lacks content validity (Chen, 2010). Well-designed multiple-choice exams may assess students' higher-order thinking abilities, such as author inference and inferred meanings (Epstein, Lazarus, Calvano, Matthews, Hendel, Epstein & Brosvic, 2002). Multiple choice testing provides a more comprehensive and extensive understanding of the topic than writing essays. Paxton (2000) argues that the multiple-choice testing technique is controversial due to poorly constructed items. One of the primary drawbacks of multiple-choice questions is their inability to assess originality. The weakness of this multiple-choice question lies in its failure to effectively integrate ideas into a new whole, as shown by the level of Bloom's taxonomy (Bloom, Engelhart, Furst, Hill & Krathwohl, 1956). Each of the five components of Bloom's taxonomy may be assessed using multiple-choice questions.

Instead of testing the concepts of higher order learning, multiple choice questions provide students access to material requiring a deeper degree of comprehension. The exam gauges the breadth and depth of the test takers' knowledge if the assessment items are properly created (Epstein et al. 2002). It is argued against the validity of multiple-choice questions (MCQs) since knowledge assessment does not imply competency. Communication skills, knowledge, attitudes, and abilities are all components of professional competence. Another disadvantage is that students might estimate the correct response without fully comprehending the reading text, raising doubts about the legitimacy of the answer (Nevo, 1989; Weir, 1993).

Subsequent research has shown that competency is contingent upon one's knowledge area. Multiple-choice questioning is reliable for assessing linguistic competency, but written tests are the only way to measure cognitive ability (Downing, 2002). The primary objective of multiple-choice questions is to assess one's knowledge. Instead of assessing the rote memorizing of isolated facts, well-crafted multiple-choice questions can potentially assess more advanced cognitive abilities such as synthesis, interpretation, and the application of information (Solano-Flores & Trumbull, 2003). MCQs are practical and effective for assessing reading comprehension since they are based on higher-order cognitive knowledge. This makes them widely accepted by both examiners and examinees. Studies have shown that individuals with lower ability levels prefer multiple choice questions (MCQs) over open-ended (Shohamy, 1984). The true/false item format is a beneficial test for evaluating reading abilities. The true/false format evaluates students'' capacity to discern the correct definition and precise factual statement. The children must respond to a range of statements by indicating whether they are true or untrue. Elementary school students' comprehension of beliefs is evaluated by requiring them to respond with either true or false.

Cause-and-effect linkages are used to assess the test takers' degree of understanding. The students are supplied with two statements and are required to ascertain the true nature of their link. One of the primary benefits of true false is its high efficiency (Rodriguez, 2005). They advocate for true/false items, asserting

that they may serve as a universal measure of linguistic knowledge. According to the recommendations of Linn & Miller (2008), there are many factors to consider when categorizing objects as either true or untrue. It is recommended to avoid making broad or unimportant statements. Instead, there should be a balance between factual and deceptive assertions.

Furthermore, it is recommended to avoid making comments that use long and pessimistic expressions. It is advisable to refrain from merging two ideas into one sentence. Reading involves the reader's deliberate attempt to understand the text to effectively convey the writer's message (Gebhard, 2006; Wallace, 1992). Hence, there is a correlation between reading behavior and purpose and a connection between reading purpose and social behaviors, which are both influenced by the situational environment and reading behavior. It was shown that multiple-choice questions (MCQs) were more effective than true/false test items. The correlation between true/false and multiple-choice questions (MCQ) was 0.872, indicating a strong positive relationship. The findings indicate that using multiple-choice questions (MCQs) in reading evaluations may significantly enhance readers' comprehension. During the act of reading, there is an interaction between the reader and the text. This interaction leads to the formation of a link between the reader's objective and their prior knowledge. Consequently, this connection influences the reader's behavior while reading.

When evaluating reading comprehension, matching items might be beneficial for measuring vocabulary within a given context. Brown (2004) states that matching words are used in examinations designed for beginner-level students. These tests require students to properly associate items, such as synonyms, antonyms, and interpreting different signals or labels. Like other techniques, matching has both benefits and downsides. According to Brown (2004), matching items provide an alternative to multiple-choice questions or fill-in-the-blank tasks, which is their primary advantage. While it is very easy to create well-designed matching objects, learners in real-world settings seldom encounter such exercises.

Picture-cued tests are used in two different variations at the first level. Upon completing a chapter, students choose the graphics that most accurately depict the content (Brown, 2004). Kitao and Kitao (1997) suggest that this exercise may be altered to assess a student's higher cognitive capacity. This can be done by asking them to choose the words that accurately describe a particular picture or series of images. Diagram labeling is a technique used in the IELTS reading examination to evaluate the level of understanding in reading. An open-ended cloze exam is an extra assessment tool to evaluate reading comprehension. Kintsch and Yarbrough (1982) conducted a study to investigate the cloze test's impact and open-ended response's impact on reading comprehension. It has been shown that whereas cloze exams primarily assess students' understanding of specific details (local meanings), they do not allow for an evaluation of their overall comprehension of the text.

On the other hand, open-ended response examinations evaluate the test takers' understanding of the main concepts in the reading material. Based on the study conducted by Pishghadam and Tabataba'ian (2011), performance is enhanced when the task is linguistically demanding due to textual structure and other signals that include several gaps. Moreover, this method has been extensively tested and is often regarded as the most effective technique to evaluate reading comprehension.

A cloze test, often used at both basic and advanced levels, evaluates textual knowledge, including global comprehension and the coherence of a book. According to Heaton (1990), summary writing focuses on developing a high-level understanding of the material. Brown (2004), Heaton (1990), and Weir (1990) saw both the advantages and disadvantages of this testing approach. For instance, a cloze test may serve as a diagnostic instrument, and it has been used by several studies in practical language training and phrase processing (Al-Shumaimeri, 2006). This test may be used to analyze phrases with a flexible framework when specific information is required or at the macro-structural level when broader information is necessary. Greene (2001) suggests that a cloze test may be used to evaluate the overall

coherence and macro level comprehension in reading comprehension. Shanahan, Kamil, and Tobin (1982) concluded that cloze tests are unsuitable for promoting overall comprehension based on their analysis of several cloze tests, including natural, intruded, and scrambled variants. According to Mckenna and Layton (1990), cloze examinations evaluate abilities at a level that goes beyond individual sentences, comparable to how TOFEL uses sentence insertion and summary completion. Hicks and Monroe (1979) provide study findings that indicate a tenuous association between casual reading materials and cloze tests. Weir (1990) states that open-ended items, in contrast to multiple-choice questions (MCQ tests), seldom provide the right answer, and guessing is nearly always prohibited. According to the findings of Brown's (2004) study, using open-ended questions stimulates students to use their language skills more imaginatively and innovatively. This might stimulate subsequent discussions and result in a positive washback effect. The main disadvantage of using this item is that the teacher may evaluate the student's writing proficiency instead of their reading comprehension. Nuttall (2005) states that the teacher intentionally formulates each question with a certain answer in mind. This prevents confusion among students and discourages them from searching for other answers.

Moreover, the process of assigning scores is fairly subjective. Evaluating somewhat accurate, wrong, and completely correct replies poses a challenge for instructors. No one testing item format comprehensively evaluates all skills and subskills simultaneously. Reading assessments use many item formats to evaluate advanced and basic cognitive skills. The GRE, TOEFL, and IELTS, prominent high-stakes testing organizations, mostly use cloze and multiple-choice question formats to evaluate reading comprehension.

BACKGROUND KNOWLEDGE

Research supports the significance of learners' previous knowledge of a topic as it enhances memory and reading comprehension. Kintsch (1988) used the constructive-integration process theory to elucidate the role of past knowledge in comprehending text. He states that a situational model is formed when the combination of informative text and previous knowledge. The situational model fully evolves as the reader becomes acquainted with the material. Eskey (1986) identified form and content as the two categories of the schema. Understanding the content of a document helps us learn about its practical, cultural, and topical features.

On the other hand, understanding the shape of a text offers us linguistic information about its semantic and syntactic patterns. Brandão and Oakhill (2005) argue that to effectively understand implicit information and produce favorable results, readers must use their prior knowledge. Studies have been conducted on the relationship between prior knowledge and reading comprehension. The participants were chosen from certain regions and societies where the researchers had previously recorded positive washback effects. Negative washback was a result of their cultural and personal life settings rather than being caused by external factors. According to Aleptekin (2006), Chen and Donin (1997), Lee (2007), and Yuet and Chan (2003), students met the selected themes in their daily lives. Previous studies have shown that a robust knowledge foundation enhances learners' ability to comprehend written content (Taft & Leslie, 1985). According to Pearson, Hansen, and Gordon (1979), Marr and Gormley (1982), and other academics, it has been shown that struggling readers often comprehend literature on a certain topic due to their substantial prior knowledge. Readers with extensive previous knowledge can construct detailed situational models and comprehend information more precisely. Two main factors contributing to this model's creation are the reader's existing knowledge and the specific needs of the reading task (Blanc & Taperio, 2001).

The main objective of reading comprehension is to combine written content with existing knowledge. Excessive usage of background and textual information might lead to a lack of comprehension and hinder the development of the situational model (Brandão & Oakhill, 2005; Canin & Oakhill, 2001; Kamalski, Sanders & Lentz, 2008; Kintsch, 1988; Pudilo, 2007). In 2005, Brandão and Oakhill conducted research

to examine the impact of prior information on the comprehension of narrative texts. The study included young children instructed to articulate their reasoning process in solving various comprehension tasks. The data revealed that both individuals used their existing knowledge and relied on information from the text. The study's findings indicate that 6.46% of students relied on their existing knowledge rather than the provided text to solve the questions. Based on the results, it was found that those students were not constructing situational models as described by Kintsch (1988). They were unable to establish a connection between new and old information. Pudilo (2007) conducted a study comparable to the research conducted by Brandão and Oakhill (2005). Pudilo concluded that background knowledge and explicit textual information work together to generate a coherent mental representation of a text.

The manner in which textual information interacts with prior knowledge significantly impacts a reader's ability to derive inferences from reading. A Previous study suggests that fourth-grade students' ability to make inferences is more influenced by their prior knowledge of the issue than their reading comprehension skills (Marr & Gormley, 1982). Occasionally, previous knowledge does not affect the process of drawing conclusions and comprehending information. However, it is the ability to integrate previously acquired knowledge with freshly presented information from the text. Lin (2002) examined the viewpoints of English as Foreign Language (EFL) learners on their past knowledge and its impact on comprehension. The study's findings indicated that most English as a Foreign Language (EFL) students associated reading comprehension with their existing knowledge. According to middle school learners, readers' prior knowledge is the most prevalent kind of background while reading English literature. According to Eskey (1986), this linguistic knowledge pertains to understanding form.

On the other hand, postsecondary students argue that sociocultural knowledge is the most prevalent kind of prior knowledge (Lin, 2002). Lin (2002) states that Garth-McCullough (2008) conducted a study investigating the relationship between reading comprehension and culture by examining the previous knowledge of African American students. Based on the research, students who possess a solid grounding in their cultural background have enhanced their ability to understand and appreciate other cultures while engaging in reading activities (Garth-McCullough, 2008). Multiple studies on the relationship between background knowledge and reading comprehension have shown that some disciplines and cultures have positive washback effects (Chen & Donin, 1997; Lee, 2007; Yuet & Chan, 2003). These courses exposed students to real-life situations based on specific cultures and specialized sectors. However, a limited number of studies included participants from other fields or cultural backgrounds. Test bias is a substantial concern that also impacts comprehension concerning prior knowledge. Therefore, it is quite challenging to develop a completely unbiased test, since bias often arises from pre-existing information. It would be unwise to do so since previous knowledge is essential for reading comprehension and it is impossible to quantify the brain's contents. Consequently, the assessment of reading is consistently dependent on inference.

CONCLUSION

This study explores several approaches to developing reading tests within the framework of authentic research and industry norms used by high-stakes testing bodies like TOEFL, GRE, and IELTS. The results showed that reading ability is assessed using different test questions, text choices, test formats, and previous knowledge, depending on the learners' cognitive level. Text selection is a crucial element of a reading comprehension test. Considerable scholarly research has focused on choosing texts according to the educational objectives of test takers, their applicability to real-life scenarios, degree of engagement, organization of the text, and cultural background. Suppose a text with a high level of linguistic complexity is selected based on cultural appropriateness and learning objectives. In that case, there is a possibility that learners may struggle to perform effectively due to language barriers. Therefore, choosing a book and categorizing it according to lower and higher cognitive capacities is more advantageous.

Sometimes, the same text is used to test many levels by adjusting the degree of difficulty. Test designers believe the text type and text selection rules established by Brown (2004) and acknowledged by international testing organizations like the GRE, TOFEL, and IELTS to be advantageous. As Brown (2004) stated, learners may understand and comprehend these suggested types of writing by using their existing knowledge. Simple narrative passages might be advantageous for elementary school students. As their vocabulary expands, individuals might use linked phrases, relative clauses, metaphors, and other linguistic devices to enhance the complexity of their writing. At the primary school level, text types may include descriptive and narrative forms like newspapers, greeting cards, and magazines. However, at the advanced level, students can select expository and argumentative text types, such as lab reports, technical writings, and professional writings, based on their interests in society and culture. Text length exhibits varying degrees of variance. The word count for texts at the basic level often ranges from 250 to 300 words, while texts at the intermediate level usually include between 300 and 800 words. On the other hand, texts at the advanced level generally consist of 1200 words or more. The length of the text might vary from short to lengthy. The test developer can include several components, and students with limited competence may get higher scores if the test creator designs an assessment consisting of a text length of 1000 words. Discourse markers and contextual signals assist in enhancing reading comprehension. Texts of shorter length indicate a lower level of proficiency since they provide fewer clues or indications.

Assessing reading abilities may be accomplished using several methods, such as multiple-choice questions (MCQs), cloze tests, fill-in-the-blank questions, and true-false questions. The true/false and multiple choice cloze tests are the most disputed methods for evaluating reading skills in the IELTS and TOEFL exams. True/false questions may be used to evaluate reading comprehension and measure the first two levels of cognition in Bloom's taxonomy. However, in practical terms, they are deemed ineffective. Researchers observed that examinees performed more effectively on multiple-choice questions (MCQs), and they identified the most significant disparities in using these formats. While several researchers ardently support multiple-choice questions, others prefer cloze tests. When Bloom's taxonomy framework is used to assess the Cloze test, it is determined that it is valid for analysis up to the fourth level. Multiple-choice (MCQ) exams are useful only until the test-taker infers the intended meanings given by the author. Researchers that prefer alternative methods are evaluating reading comprehension using summary rather than the traditional exam. Multiple-choice question (MCQ) examinations are more valid than cloze tests in this context. Contemporary testing companies use a variety of test formats in the present day. Assessing test takers' abilities using a single-item format makes it challenging to conduct comprehensive testing. A reading assessment is a thorough examination of an individual's abilities. Consequently, the testing's goals should dictate the structure and selection of material.

PEDAGOGICAL IMPLICATIONS

Different texts should be utilized depending on the level. Different genres are used for different exam goals throughout the selection process. Primary school students are suitable for short stories and other narrative works. At the intermediate level, descriptive writing encompasses several tasks, such as technical writing, map interpretation, letter and email composition, and historical event description. Advanced-level choices include a range of possibilities, such as reports, instructional manuals, immigration papers, and expository and argumentative essays. The required word count for texts at the elementary, intermediate, and advanced levels should be 250–300, 400–850, and beyond 1200 words, respectively. Novices should be given test items in the form of true-false statements, matching exercises, and visual prompts to assess their level of comprehension and knowledge. At the intermediate level, activities such as summary cloze, cloze elide, and information transfer may be provided. Advanced level assessments should have multiple-choice questions that require students to infer meanings from authors

since this will effectively evaluate their higher cognitive talents. Test takers will need to infer meanings from the questions.

REFERENCES

Alderson, J. C. (1990). Testing reading comprehension skills (Part one). *Reading in a Foreign Language*, 6(2), 425-438.

Alderson, J. C. (2000). *Assessing reading*. Cambridge, UK: Cambridge University Press.

Alderson, J. C., & Lukmani, Y. (1989). Cognition and reading: Cognitive levels as embodied in test questions. *Reading in a Foreign Language*, 5(2), 253-270.

Ali, Z., Palpanadan, S. T., Asad, M. M., Churi, P., & Namaziandost, E. (2022). Reading approaches practiced in EFL classrooms: a narrative review and research agenda. *Asian-Pacific Journal of Second and Foreign Language Education*, *7*(1), 28.

Alptekin, C. (2006). Cultural familiarity in inferential and literal comprehension in L2 reading. *System*, 34(4), 494-508.

Al-Shumaimeri, Y. (2006). The effects of content familiarity and language ability on reading comprehension performance of low and high-ability Saudi tertiary students studying English as a foreign language. *Education Science and Islamic Studies*, 18, 1-19.

Anaya, L., Iriberri, N., Rey-Biel, P., & Zamarro, G. (2022). Understanding performance in test taking: The role of question difficulty order. *Economics of Education Review*, *90*, 102293.

Andreassen, R., & Bråten, I. (2010). Examining the prediction of reading comprehension on different multiple-choice tests. *Journal of Research in Reading*, 33(3), 263-283.

Arias, I. J. (2007). Selecting reading materials wisely. *Letras*, 41, 131-151.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press.

Berkoff, N. A. (1979). Reading skills in extended discourse in English as a foreign language. *Journal of Research in Reading*, 2(2), 95-107.

Bishop-Monroe, R., Jordan, M., Ma, Z., & Royalty, K. (2022). Enhancing business professional competencies in a virtual educational environment. *The International Journal of Management Education*, *20*(3), 100700.

Blanc, N., & Tapiero, I. (2001). Updating spatial situational models: Effects of prior knowledge and task demands. *Discourse Processes*, 31(3), 241-262.

Blaži Ostojić, A. (2023). Reading comprehension processes: a review based on theoretical models and research methodology. *Hrvatska revija za rehabilitacijska istraživanja*, *59*(1), 122-143.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, NY: David McKay.

Brandão, A. C. P., & Oakhill, J. (2005). "How do you know this answer?" – Children's use of text data and general knowledge in story comprehension. *Reading and Writing*, 18(7-9), 687-713.

Brown, D. H. (2004). *Language assessment: Principles and classroom practices*. New York, NY: Longman.

Brysbaert, M. (2022). Word recognition II: phonological coding in reading. *The science of reading: A handbook*, 79-101.

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition*, 29(7), 850-859.

Carter, M., Rastatter, M. P., Walker, M. M., & O'Brien, K. (2009). The effects of frequency altered feedback on the reading processes of adults with reading disorders. *Neuroscience Letters*, 461(2), 69-73.

Carver, R. P. (1990). *Reading rate: A review of research and theory*. New York, NY: Academic Press.

Catts, H. W. (2022). Rethinking How to Promote Reading Comprehension. *American Educator*, *45*(4), 26.

Chan, N., & Kennedy, P. E. (2002). Are multiple-choice exams easier for economics students? A comparison of multiple-choice and "equivalent" constructed-response exam questions. *Southern Economic Journal*, 68(4), 957-971.

Chen, C. (2010). On reading test and its validity. *Asian Social Science*, 6(12), 192-195.

Chen, Q., & Donin, J. (1997). Discourse processing of first and second language biology texts: Effects of language proficiency and domain-specific knowledge. *The Modern Language Journal*, 81(2), 209-227.

Conrad, C., Deng, Q., Caron, I., Shkurska, O., Skerrett, P., & Sundararajan, B. (2022). How student perceptions about online learning difficulty influenced their satisfaction during Canada's Covid-19 response. *British Journal of Educational Technology*, *53*(3), 534-557.

Conradi Smith, K., Young, C. A., & Core Yatzeck, J. (2022). What are teachers reading and why?: An analysis of elementary read aloud titles and the rationales underlying teachers' selections. *Literacy Research and Instruction*, *61*(4), 383-401.

Cutting, L. E. (2017). Do all reading comprehension tests assess equally? [Blog post]. Houghton Mifflin Harcourt. Retrieved April 11, 2020, from https://www.hmhco.com/blog/do-all-reading-comprehension-tests-assess-equally

Davis, F. B. (1968). Research in comprehension in reading. *Reading Research Quarterly*, 3(4), 499-545.

Day, R. R. (1994). Selecting a passage for the EFL reading class. *English Teaching Forum*, 32(1), 20-23.

Day, R. R., & Park, J. S. (2005). Developing reading comprehension questions. *Reading in a Foreign Language*, 17(1), 60-73.

Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., & Brosvic, G. M. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record,* 52(2), 187-201.

Eskey, D. E. (1986). Theoretical foundations. In F. Dublin, D. E. Eskey, & W. Grabe (Eds.), Teaching second language reading for academic purposes (pp. 3-23). *Reading,* MA: Addison-Wesley.

Feng, R., Alsager, H. N., Azizi, Z., & Sarabani, L. (2023). Impact of mind-mapping technique on EFL learners' vocabulary recall and retention, learning motivation, and willingness to communicate. *Heliyon*, *9*(6).

Ferris, D. R., & Hedgcock, J. S. (2023). *Teaching L2 composition: Purpose, process, and practice*. Routledge.

Gallagher, K. (2023). *Deeper reading: Comprehending challenging texts, 4-12*. Routledge.

Garth-McCullough, R. (2008). Untapped cultural support: The influence of culturally bound prior knowledge on comprehension performance. *Reading Horizons: A Journal of Literacy and Language Arts,* 49(1), 1-30.

Gebhard, J. G. (2006). Teaching English as a foreign or second language: A teacher self-development and methodology guide. Ann Arbor, MI: University of Michigan Press.

Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly*, 25(3), 375-406.

Grabe, W., & Stoller, F. (2002). *Teaching and researching reading.* New York, NY: Longman.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items.* London, United Kingdom: Routledge.

Han, C. (2022). Interpreting testing and assessment: A state-of-the-art review. *Language Testing*, *39*(1), 30-55.

Hattan, C., Alexander, P. A., & Lupo, S. M. (2024). Leveraging what students know to make sense of texts: What the research says about prior knowledge activation. *Review of Educational Research*, *94*(1), 73-111.

Heaton, J. B. (1975). *Writing English language tests: A practical guide for teachers of English as a second or foreign language*. London, United Kingdom: Longman.

Heaton, J. B. (1988). *Writing English language tests.* London, United Kingdom: Longman.

Heaton, J. B. (1990). *Classroom testing.* London, United Kingdom: Longman.

Hicks, R. D., & Monroe, E. (1979). A comparison of reading achievement, current reading placement, sex, age, intelligence, informal reading inventory, and the cloze procedure. *ERIC* Document Reproduction Service No. ED181438.

Hudson, T. (2007). *Teaching second language reading.* Oxford, United Kingdom: Oxford University Press.

Hughes, A. (2003). *Testing for language teachers (2nd ed.).* Cambridge, United Kingdom: Cambridge University Press.

Jian, Y. C. (2022). Reading in print versus digital media uses different cognitive strategies: evidence from eye movements during science-text reading. *Reading and Writing*, *35*(7), 1549-1568.

Johnson, A. P., & Lester, R. J. (2022). Mental health in academia: Hacks for cultivating and sustaining wellbeing. *American Journal of Human Biology*, *34*, e23664.

Kamalski, J., Sanders, T., & Lentz, L. (2008). Coherence marking, prior knowledge, and comprehension of informative and persuasive texts: Sorting things out. *Discourse Processes*, 45(4-5), 323-345.

Kim, A. Y. (2009). Investigating second language reading components: Reading for different types of meaning.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review,* 95(2), 163-182.

Kintsch, W., & Yarbrough, J. C. (1982). Role of rhetorical structure in text comprehension. *Journal of Educational Psychology,* 74(6), 828-834.

Kitao, K., & Kitao, S. K. (1997). Selecting and developing teaching/learning materials. *The Internet TESL Journal,* 4(4), 20-45.

Kobayashi, M. (2002). Method effects on reading comprehension test performance: Text organization and response format. *Language Testing,* 19(2), 193-220.

Kobayashi, M. (2005). An investigation of method effects on reading comprehension test performance. *Paper presented at the Proceedings of the 4th Annual JALT Pan-SIG Conference, Tokyo, Japan.*

Kuromiya, H., Majumdar, R., Miyabe, G., & Ogata, H. (2022). E-book-based learning activity during COVID-19: engagement behaviors and perceptions of Japanese junior-high school students. *Research and Practice in Technology Enhanced Learning*, *17*(1), 12.

Lee, S. K. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning,* 57(1), 87-118.

Lin, Z. (2002). Discovering EFL learners' perception of prior knowledge and its roles in reading comprehension. *Journal of Research in Reading,* 25(2), 172-190.

Linn, R. L., & Miller, M. D. (2008). *Measurement and assessment in teaching.* New Delhi, India: Pearson Education.

Liu, F. (2011). A short analysis of the text variables affecting reading and testing reading. *Studies in Literature and Language,* 2(2), 44-49.

Lumley, T. (1993). The notion of subskills in reading comprehension tests: An EAP example. *Language Testing,* 10(3), 211-234.

Marr, M. B., & Gormley, K. (1982). Children's recall of familiar and unfamiliar text. *Reading Research Quarterly,* 18(1), 89-104.

Marzano, R., & Kendall, J. (2007). *The new taxonomy of educational objectives (2nd ed.).* Thousand Oaks, CA: Corwin Press.

Matthews, M. (1990). Skill taxonomies and problems for the testing of reading. *Reading in a Foreign Language,* 7(1), 511-517.

McKenna, M. C., & Layton, K. (1990). Concurrent validity of cloze as a measure of intersentential comprehension. *Journal of Educational Psychology,* 82(2), 372-377.

McKenna, M. C., & Stahl, K. (2015). Assessment for reading instruction. New York, NY: *Guilford Press.*

McNamara, T., & Roever, C. (2006). Language testing: The social dimension. *Malden, MA: Blackwell.*

Munby, J. (1978). *Communicative syllabus design.* Cambridge, United Kingdom: Cambridge University Press.

Nevo, N. (1989). Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing*, 6(2), 199-215.

Nuttall, C. (1996). *Teaching reading skills in a foreign language.* Oxford, United Kingdom: Heinemann International.

Oktarina, Y., Inderawati, R., & Petrus, I. (2022). Needs analysis of Palembang-tourist-destination recount text reading materials in the 21st century learning. *English Review: Journal of English Education*, *10*(2), 381-392.

Osterlind, S. J. (1998). *Constructing test items: Multiple-choice, constructed-response, performance and other formats (2nd ed.).* New York, NY: Kluwer Academic Publishers.

Paxton, M. (2000). A linguistic perspective on multiple choice questioning. *Assessment & Evaluation in Higher Education,* 25(2), 109-119.

Pearson, P. D., & Johnson, D. D. (1978). *Teaching reading comprehension.* New York, NY: Holt, Rinehart and Winston.

Pearson, P. D., Hansen, J., & Gordon, C. (1979). The effect of background knowledge on young children's comprehension of explicit and implicit information. *Journal of Reading Behavior,* 11(3), 201-209.

Perfetti, C., & Helder, A. (2022). Progress in reading science: Word identification, comprehension, and universal perspectives. *The science of reading: A handbook*, 5-35.

Pham, V. P. H., & Bui, T. K. (2022). Genre-based approach to writing in EFL contexts. *Pham, VPH, & Bui, TKL (2021). Genre-based Approach to Writing in EFL Contexts. World Journal of English Language*, *11*(2), 95-106.

Pishghadam, R., & Tabatabaian, M. (2011). IQ and test format: A study into test fairness. *Iranian Journal of Language Testing,* 1(1), 17-29.

Pulido, D. (2007). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning,* 57, 155-199.

Rezaei Ghahroudi, M., & Sheikhzadeh, E. (2017). Selecting reading texts for university Iranian EFL students. *Journal of Advances in English Language Teaching,* 5(3), 25-30.

Rost, D. H. (1993). Assessing different components of reading comprehension: Fact or fiction? *Journal of Educational Measurement,* 27(2), 109-131.

Sainsbury, M., Harrison, C., & Watts, A. (2006). *Assessing reading: From theories to classrooms.* Slough, United Kingdom: National Foundation for Educational Research.

Shah, S. H. R., Kadir, Z. A., & Naveed, S. (2022). Factors affecting English reading skills at the collegiate level in Pakistan. *Journal of Positive School Psychology*, 6(11), 1863-1876.

Shah, S. H.R., Altaf, A. R., Mughal, K. H. (2025). Exploring the Impact of AI Tools on Reading Proficiency Among Undergraduate Students. *The Critical Review of Social Sciences Studies*, *3*(3), 663-670.

Sherkuziyeva, N., Imamutdinovna Gabidullina, F., Ahmed Abdel-Al Ibrahim, K., & Bayat, S. (2023). The comparative effect of computerized dynamic assessment and rater mediated assessment on EFL learners' oral proficiency, writing performance, and test anxiety. *Language Testing in Asia*, *13*(1), 15.

Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing,* 147(1), 147-170.

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher,* 32(2), 3-13.

Taft, M. L., & Leslie, L. (1985). The effects of prior knowledge and oral reading accuracy on miscues and comprehension. *Journal of Reading Behavior,* 17(2), 163-179.

Urquhart, A. (1998). *Reading in second language: Process, product and practice.* New York, NY: Longman.

Wallace, C. (1992). *Reading.* New York, NY: Oxford University Press.

Weir, C. J. (1990). *Communicative language testing. London,* United Kingdom: Prentice Hall.

Weir, C. J. (1993). *Understanding and developing language tests.* New York, NY: Prentice Hall.

Weir, C. J., & Porter, D. (1994). The multi-divisible or unitary nature of reading: The language tester between Scylla and Charybdis. *Reading in a Foreign Language,* 10(2), 1-19.

Wolfe, M. B., & Woodwyk, J. M. (2010). Processing and memory of information presented in narrative or expository texts. *British Journal of Educational Psychology,* 80(3), 341-362.

Wyse, D., & Bradbury, A. (2022). Reading wars or reading reconciliation? A critical examination of robust research evidence, curriculum policy and teachers' practices for teaching phonics and reading. *Review of education*, *10*(1), e3314.

Yuet, C., & Chan, H. (2003). Cultural content and reading proficiency: A comparison of mainland Chinese and Hong Kong learners of English. *Language Culture and Curriculum,* 16(1), 60-69.

Zheng, Y., Cheng, L., & Klinger, D. A. (2007). Do test formats in reading comprehension affect second-language students' test performance differently? *TESL Canada Journal,* 65-80.