A Review on the Principles of a Reading Comprehension Test Construction to Assess the Test Takers at Different Levels

Amna Arshad
PhD Candidate, Department of Applied Linguistics, Government College University, Faisalabad, Pakistan
Dr. Aleem Shakir
Assistant Professor, Department of Applied Linguistics, Government College University, Faisalabad, Pakistan
Muhammad Ahmad (Corresponding Author)
PhD Candidate, Department of Applied Linguistics, Government College University, Faisalabad, Pakistan
Email: ahmad453@yandex.com

ABSTRACT
The aim of this review article is to explore the principles involved in testing reading abilities. It considers the key issues i.e. the selection of text for testing reading comprehension, item format, text length, use of different genres and the role of background knowledge to test individual student levels or processes. Most of the studies followed different taxonomies to test reading comprehension. Bloom's taxonomy had extensively been used by many researchers to measure low and high order skills. The data was collected from different sources i.e. books and journals. Results, based on previous literature, showed that the construction of reading tests was highly dependent on the purpose of measuring intended skills or sub skills at lower and higher cognitive levels of the students. Text selection and its length, text type, item formats, background knowledge of test takers were considered according to the levels of test takers.

KEYWORDS: *background knowledge; Bloom's taxonomy; reading comprehension tests; test formats; text length; text selection*
Article Received: 10 August 2020, Revised: 25 October 2020, Accepted: 18 November 2020

Introduction

Testing is basically a procedure for measuring knowledge, ability or performance of a test taker. Brown (2004) defines testing as a method of measuring individual's knowledge or ability in a given domain. The term method refers to a set of procedures, techniques and items that demands some task to perform on the part of test taker by means of an instrument. Heaton (2003) defines tests as an instrument for assessing students' performance in language that reinforces learning with motivation among students. Tests also assess candidates on the basis of relevant and purposeful tasks that are actually used in their real life communication by performing relevant tasks. Therefore, it is inferred that testing instruments measure knowledge and ability to perform linguistic tasks in actual life. In testing reading comprehension, reading can be seen as product and process (Alderson, 2000). Process is an interaction between text and the reader. Product refers to the results of the process. This seems to be meant that test a design of comprehending a particular text, suitable test item formats to administer a test, using a specific research design and then examining the results. Since student may have different knowledge and life experiences, then the product of reading will be different necessarily.

Reading as one of the critical skills in academics, is used to acquire new information, and labeled as one of the essential skills in English as a second or foreign language. Many high stakes exams (e.g. IELTS, TOEFL, TOEIC) give high importance to reading comprehension ability. Reading is an interactive process that includes activation of different abilities simultaneously. The interaction also involves higher level comprehension, and it consists of the relationship between reader's prior knowledge and information possessed by the text.

Moreover, reading comprehension includes both types of information i.e., information printed in the form of a plain text and information that is interpreted by the readers and

when they relate to a new set of information with their prior knowledge. Although, meanings reside within text, readers from different cultures, knowledge and interests interpret text through different ways and sometimes their interpretations are different by reading the same text in different context. Significant number of researches has been conducted to investigate the nature of reading skills. However, defining reading skill development, teaching and its assessment, is a difficult task and varies from context to context (Hedgcock & Ferris, 2009; Sainsbury, Harrison & Watts, 2006).

Mainly there are two contrastive views on reading comprehension skills i.e. reading as a unitary skill, and reading as a multi-divisible skill (Weir & Porter, 1994). The former view implies that reading abilities have not separate sub-components or sub skills (Alderson, 1990; Bachman, 1990; Rost, 1993; Weir & Porter, 1994).

Basically, reading is characterized by two integrated components namely: word recognition (decoding); and comprehension. Hughes (2003) explains issues regarding sub skills in developing reading skills, while in testing reading comprehension; it must include skills that are intended to be tested. He further maintains that testing reading skills should be followed by clear theory of reading that clearly defines the reading abilities, and helps measure and interpret student's performance accurately (Engelhard, 2001).

The later view of reading (multi-divisibility) explicates that a particular skill of reading has identifiable and separate sub skills. As it is evidenced greatly in the previous studies that one particular reading skill can be divided into different sub skills (Davis, 1968; Watts et al. 2006; Kim, 2009; Matthews, 1990; Munby, 1978). In this respect, the relationship between high and low order thinking skills in testing reading and distinction between them is necessary to understand better about the nature of reading abilities, moreover, for valid test item development to assess reading abilities and also for syllabus planning (Hughes, 2003; Lumley, 1993; Weir & Porter, 1994). Multi-divisible view regarding reading skills clouds with important issue i.e., the ordering of skills or sub skills of testing reading and testing number of abilities. Researches in first language and second language have failed in positioning hierarchical order of reading skills. There is no clear cut evidence for the strict division to test high and low order reading skills (Alderson, 1990; Hudson, 2007).

Despite of all the facts, the views in the favour of skills and sub skills of reading are influential. This can be seen in the use of different taxonomies utilized in teaching and testing of reading (Alderson & Lukmani, 1989; Grabe, 1991; Hudson, 2007; Matthews, 1990; Pearson & Johnson, 1978; Urquhart, 1998; Vacca & Vacca, 2005). In spite of the utility of all these taxonomies, these are also criticized, for instance, Matthews (1990) states that Munby's taxonomy (1978) is based on knowledge rather skills. Hudson (2007) argues that Bloom's, Barrett's and Munby's taxonomies are largely based on theoretical assumptions rather than empirical frameworks.

Different factors are involved that influence the testing reading skills like textual information, test items, and knowledge. In reading testing, item difficulty is influenced by a number of factors rather than inherent skill difficulty (Day & Park, 2005; Kobayashi, 2005). These factors involve item format, context type, implicitness and explicitness of information, test taking strategies of students' and demand of higher cognitive level (Mckenna & Stahl, 2015; Pearson & Johnson, 1978), text length and text type (Scheuneman & Gerritz, 1990). With respect to the item format, multiple choice items are influenced by its stem content, length, options, length of correct option and distracters. Pearson and Johnson (1978) find that implicit or explicit information, provided in questions, also influences the item difficulty. Hence, for valid conclusions of high and low order skills, it is important to consider these factors that are influential on item difficulty. Kintsch (1988) uses theory of constructive-integrative process which states the positive role of background knowledge in testing reading comprehension.

This study is a review article on the principles that help construct reading tests. Two types of variables are included in testing reading i.e. textual variables which consist of text length, selection of text, genre, commonly used item formats, and reader variables in which different levels of learners, schema, and their culture are considered while constructing reading comprehension test. Bloom's taxonomy as a practical basis for assessment (Marzano & Kendall, 2007) is helpful to measure different abilities of learners by analyzing different types of testing items with respect to the level of learners. This study helps future test constructors to test reading abilities by considering the discussed principles and goal of testing could be achieved in a valid manner. Thus, realizing the importance of reading comprehension, this review extracts principles of testing reading comprehension (i.e. text selection principles, text length principles, principles of constructing item formats for

testing reading at different levels, background knowledge, and type of skills to be measured in a reading comprehension test) by reviewing literature from authentic researches and books.

1.1 Text Selection

Text selection is an important part for a constructing reading test. Previous studies show that there are two types of criteria on which selection is made i.e.: criteria related with students; and criteria related with text selection. In students related criteria, factors like students' interests, level of content, needs, purpose of reading, and cultural schema are needed to be catered (Arias, 2007). According to Berkoff (1979), a test constructor should select the text according to students' linguistic level (i.e. grammar, vocabulary and syntactic complexity), cultural schema and their content level. Shoebottom in Rezaei Ghahroudi and Sheikhzadeh (2017) pointed out that the selection of a text specifies learners' reading purpose. In this way, skills and sub skills of reading can be tested. The prior knowledge of students should also be considered and provide students with necessary and appropriate background information before they are required to read the testing text (Norazit & Khemlani in Rezaei Ghahroudi & Sheikhzadeh, 2017).

A variety of test items are used in order to evaluate reading comprehension. Activities include pre-reading, skimming, scanning, reading comprehension, translation and reading aloud, reading literature etc. However, reading comprehension is extensively used in schools. Walter (1992) says that reading tests function as giving feedback (recreating text in reader's mind), since it includes responding to a particular text. Certain factors are related with the text while selecting a text e.g. the content of the text, suitability, readability, topic, appearance, political and cultural suitability (Hetherington, 1985). Arias (2007) finds that material selection should be based on students' professional settings and related with their real world situations. The selected content for testing should be based on teacher's objectives of a reading lesson. Due to lexical and thematic content it can be exploitable. Nuttall (1996) states that a text should be fit, match and delightful for the test taker. There should be appropriateness, logical and rhetorical order of ideas in the text.

Textual phenomena at discourse level and background knowledge of the reader should be valued (Day, 1994). The text should be coherent and clear to the reader (Nuttall, 1996) and the reader should not be confused by the linguistic features of different genres (Grabe & Stoller, 2002). Furthermore, Guler maintains that cultural issues in the text should not be taken for granted (2007), rather it should be according to the student's prior cultural experience (Erten & Razi, 2009).

1.2 Genre

Genre is one of the important elements in the construction of tests. Text selection largely depends on the genre. Different types of genres are used at different levels. Hughes (2003)

recommended that a reading text should provide realistic purpose of language assessment and motivate learners to use the language in a target context. Text types suggested by Brown (2004) that include; textbooks, newspaper articles, magazines, journals, and verses extracted from different poems, dictionary entries, encyclopedic entries, leaflets, letters, time tables, extracts from novels, advertisements, diaries, postcards, manual computer help systems, reviews, maps and short stories. He states that such text type readers to apply certain schemata that will assist them in extracting appropriate meaning".

Text type and genre also play an important role in testing reading comprehension. Texts are categorized into descriptive, expository, argumentative and narrative. Research has proved that expository type of text is hard to process whereas narrative texts are digestible due to relationships among textual units and rich content. The macro-structure within the stories seems to facilitate reading comprehension. Weir (1990) suggests that each text is predetermined for different methods of test and Alderson (2005) is the proponent of the same opinion and states: there is both one single method that can satisfactorily serve the purpose for all tests or for all texts and skills intended to be tested (Nuttall 2005). Other ELT experts have identified that content the affects performance of test takers. Okumura (1998) conducted a study. He found that Japanese high students were given one typical expository text and two narrative readings. He also found a significant difference between the reading score of narrative and expository texts. The students scored high in narrative text, therefore, it was noted that L2 learners' ability of reading comprehension varied according to text type. Text types are different from one another in terms of text organization. How paragraphs relate to one another and how ideas are interconnected with series of events- has been interest of researchers. Researchers considered different organizational patterns within one genre that might lead to different outcomes. Texts are categorized into four types i.e. argumentative, descriptive, narrative and expository.

Different texts types demand the reader to use different skills and their level of comprehension can be affected. Hughes (2003) enlisted different types of text e.g.: descriptive texts include technical and impressionistic descriptions; expository texts include explications, impressionistic descriptions and outlines, summaries, and text interpretations; argumentative include instructional texts, personal instructions, comments and formal argumentations, practical instructions; and narrative texts include jokes, reports, biographical notes, stories, news, and historical accounts. In testing of reading, item difficulty was found to be influenced by a number of factors rather than the inherent difficulty of the skill (Bachman, 1990; Day & Park, 2005; Kobayashi, 2005; McNamara & Roever, 2006) and those factors includes; context type, question type, question format, explicit and implicit of information, cognitive demand, students' test-taking strategies (McKenna & Stahl, 2009; Pearson & Johnson, 1978), and text type as well as text length (Scheuneman & Gerritz, 1990).

With regard to the type of question format, a multiple choice question is influenced by its stem length, stem content words, structure of options, length of correct answer and distracters. Furthermore, Pearson and Johnson (1978) find that an item's difficulty characterizes the implicitness and explicitness of question information. Brown (2004) enlists and emphasizes textual genres as it forms the specifications for testing reading ability. He divides different types of written texts of reading into three groups i.e.: academic genres (e.g. general types of interest articles like newspaper and magazines, technical reports like lab reports, articles written by professionals, reference material like dictionaries, texts books , essays, papers); and job- related readings (e.g. letters/mails, messages, reports, directions, financial documents); personal reading (e.g. novels, jokes, drama, poetry, newspaper and magazines, maps, recipes, greeting cards, invitations, messages, notes, lists, questionnaire, forms, immigration documents). All of these genres set their own governing conventions and rules for the manifestation of the text and the tester must anticipate those conventions to focus on the text and process the meanings within the text. People tend to think in quite similar ways. Therefore, readers are able to predict according to what we can expect from the text. International examining bodies e.g. IELTS selects reading texts from descriptive, narrative and argumentative tests, TOFEL tests include text from different academic writing topics e.g. art, science and psychology and GRE tests use materials from biographies, books, and textbooks. Liu (2011) has stated that many researchers have established that expository texts are harder to process than narrative texts perhaps because of diverse variety of content and due to great relationships among textual units.

1.3 Text Length

With respect to the relationship between reading comprehension ability and text length, Alderson (2000) asserts that all test-developers face problems while deciding how long the texts length should be for different levels on which their test will base. According to him, the text length is still an under-researched area in testing reading ability. It is an intuitive belief that longer the passage, the difficult it will be. However, learners' performance is determined by his/her reading ability and test method facets. Test method facets include rubrics, testing environment and input (Bachman, 1990) whereas, input format of tests encompasses text length, form, channel, degree of speediness (Bachman & Palmer, 1996). Input, present in smaller chunks, requires a limited set of interpretation whereas input which is presented in an extended discourse requires larger interpretation. Alderson (2000) maintains that the texts may be comprised of a single word, or a phrase (e.g. public notices and warnings), single sentences (e.g. advertisements), whole novel, two-paragraphs on post cards, 20-pages academic article. He further states text length is clearly related to the amount of interpretation required. Surprisingly, there is no direct relation between length and interpretation e.g. a post card can be very difficult to interpret and the novel can be explicit (Alderson, 2000). Heaton (1975) recommended 50 to 100, 200 to 300, and 400 to 600 words for primary, elementary, and advance levels respectively. Text length depends on genre from where the text is chosen. The appropriate length of a chosen text comprises of 2000 words and the passages can be selected from a series of events. According to Flesch-Kincaid index a passage length, ranging 800-1200 words, is appropriate for advance level. Alderson (2000) observed an evident change in measuring abilities of test takers when passage length was more than 1,000 words. He suggested that lengthy texts assessed student's study skills and student became less relied on inter and intra sentential abilities. Therefore, it leads to students less concentration on structural and lexical knowledge and more concentration on processing of discourse knowledge. Moreover, ability to identify the main idea in longer texts as compared to shorter texts might be different qualitatively. It is much

easier to measure reading speed by using longer texts rather than using shorter texts with associated questions. On the contrary, Andreassen and Braten (2010) stated that lengthy text might increase the demands of working memory. It has also been demonstrated that lengthy texts may lead to mind wandering happenings which affect negatively on reading comprehension. Carver (1990) recommended that text length should be determined by considering a single word as six characters including spaces was adopted. The relatively shorter texts were 300–350 words long and the relatively longer texts were 600–650 words long. Carter, Rastatter, Walker and O'Brien (2009) found that the lengthy text enhanced reading comprehension ability in testing rather than using shorter texts. Texts length was approximately 1¾, pages typed with double space by using 12 Times New Roman font. Findings showed that through accelerated reading technique longer texts not only improved reading comprehensive but also had a positive side effect. However, the current practices at international level exam e.g. IELTS uses 3 paragraphs, GRE uses 1-5 paragraphs and TOFEL uses 3-4 paragraphs. Therefore, 3-5 paragraphs can be considered a justified text length according to different levels. Wolfe and Woodwyk (2010) showed that shorter texts favoured bottom-up reading and greater attention was needed for word-by-word processing. In this way, short texts are more difficult due to limited amount of information. While lengthy texts favor top-down that are easy to understand as they provide more content and are meaningful.

In the light of previous studies, the number of shorter texts can be allowed to use for testing reading i.e. 3-5 paragraphs. However, this also allows covering a wider range of topics and test constructors' biasness can be controlled. The average text length for a shorter passage can be 300-600 words and 600-1200 words for the longer texts. Longer texts are more favorable for testing reading as a number of meta-cognitive skills includes interpreting main idea, drawing conclusions, inferring meanings from context and suitable for advance learners. No doubt, a longer text facilitates the testers through providing more information and contextual clues. In designing reading tests, test constructor should compromise by maximizing the authenticity of longer texts, which students have read in their courses earlier, and by minimizing the density of the content by using a number of shorter texts.

1.4 Types of Test Formats

A number of testing techniques for testing reading comprehension like cloze test, multiple choice questions, gap-filling, picture-cued items, skimming and scanning are included. Heaton (1990) proposed: sentence matching, word matching, pictures and sentence matching for elementary level of testing reading; and matching tests for intermediate level. For an advanced level multiple choice items, true/false reading tests, rearrangement, completion, open-ended and miscellaneous items, cloze procedure, and cursory reading are recommended. These testing techniques are used for testing low order and high order skills at different levels. Following Brown (2004) proposed written response, reading aloud, multiple choice, matching test, picture-cued items, editing, cloze test, gap filling test, C-test, cloze-elide test, ordering test, short-answer test, and summarizing test. Alderson (2000) and Cutting (2017) have suggested multiple-choice, gap-filling test, cloze test, matching, editing, ordering, and cloze-elide, summary, short-answer, gapped summary, and information-transfer.

One of the fundamental principles of an item format choice should be to measure the content and the cognitive process (Haladyna, 2004). Hughes (2003) stated that the type of task should be according to the skills which intended to test. Kobayashi (2002) noted the relationship between students test performance and two other variables such as test type and test items. In past researches mostly three test formats were utilized i.e. cloze, summary writing and open-ended questions. The findings of the study indicate that student's performances are highly correlated with both variables. In Zheng, Cheng and Klinger (2007), findings show that different types of test formats measure different types of testing abilities of learners. There is no one particular test format that assesses all the concerned abilities.

MCQs are considered as one of the most important testing format in assessing reading. As Brown (2004) states that the advantage of using this test is easy to score and administer. The most versatile available type of testing format is MCQ. From simple to complex, a variety of learning outcomes can be measured. Due to its wide applicability and adaptability to all types of subject-matter that high stakes tests use exclusively is MCQs format. A standard MCQ has three parts: A stem (states problem), the best or correct answer; and number of distracters (less appropriate and wrong). MCQ items are constructed to assess a variety of learning outcomes, from recalling facts to highest level of cognition of Bloom's taxonomy-evaluation (Osterlind, 1998). Heaton (1988) has declared that the chief objection against multiple-choice is that, it is not a test of language as communication. However, it is

also true that an MCQ test is easy to score with objectivity and accuracy. These types of tests are beneficial for poor readers and they do not place them behind like essay writing tests (Chan & Kennedy, 2002).

A major concern regarding multiple choice questions is that, these tests are subjectively designed by test constructor that can cause a lack in content validity (Chen, 2010). Well-constructed multiple choice tests can assess students' higher cognition level like inference and implied meanings of an author (Epstein, Lazarus, Calvano, Matthews, Hendel, Epstein & Brosvic, 2002). Multiple choice testing provides an increase in depth and breadth of the material coverage as compared to essay writing. However, Paxton (2000) states that multiple choices' testing technique appears to be controversial due to poorly constructed items. One of the major drawbacks of multiple choice items is that it cannot measure creativity. This weak point of multiple choice items reflects Bloom's taxonomy level of combining ideas into a new whole (Bloom, Engelhart, Furst, Hill & Krathwohl, 1956). All of the five components of Bloom's taxonomy can be tested through multiple choice questions.

Multiple choice questions provide information of higher level of understanding for the students rather than it just tests ideals of higher order learning. If testing items are designed correctly, the test measures the depth and breadth of the test takers' knowledge (Epstein et al. 2002). Criticism on the validity of MCQ items is that testing of knowledge does not guarantee ones competence. Professional competence combines skills, attitudes, and knowledge and communication skills. Another drawback is that the student can guess the right the answer without understanding the full reading passage and thus validity is questionable (Nevo, 1989; Weir, 1993).

After decades, researches showed that the domain of knowledge is the determinant of the expertise. MCQ testing is, therefore, a legitimate method of testing linguistic competence, however, cognitive competence can only be assessed by using written tests (Downing, 2002). Multiple choice questions are designed to assess knowledge. Well-constructed multiple choice questions can assess high order cognitive abilities taxonomically like synthesis, interpretation, and knowledge application rather than testing recalling isolated facts (Solano-Flores & Trumbull, 2003). MCQs, based on higher cognitive knowledge, are still applicable to testing reading comprehension and their practicality makes it more acceptable to examinees and examiners. Studies have showed that testers with low proficiency level respond better on an MCQ rather open ended or cloze items (Shohamy, 1984). Another useful test format in reading assessment is true false item format. True/false format only measures learners' ability to choose correct factual statement, and selects appropriate definition. Different statements are given to which the students are supposed to respond in true or false. Abilities of knowledge concerning beliefs are measured by giving true false at elementary level.

The aspect of understanding in test takers is measured by cause-effect relation. The students are given two propositions from which they have to judge the true relationship between them. A major benefit of using true false is that they are efficient (Rodriguez, 2005). They are the proponents of using true false items as they argue that verbal knowledge can be broadly measured by the true false. There are some considerations (as suggested by Linn & Miller, 2008) while organizing items of true and false. First of all it is suggested avoid too broad and trivial statements, the number of true statements should be equal to false statements. Secondly, it is suggested to avoid negative and long sentences in statements. Infusing two ideas in one statement should also be avoided. By reading, the reader attempts to understand the text having some purpose which the writer has intended to communicate (Gebhard, 2006; Wallace, 1992). This means that reading behavior is linked with reading purpose which is directly linked with their social practices and situational context. They also found that MCQ was more effective than using test items of true and false. The correlation between true false and MCQ was 0.872 which was significantly high. The findings show that a reader's in testing at reading comprehension, by using MCQ students' gained more comprehension in effective manner. Reading is an interaction between the text and reader that builds a link between prior knowledge and reading purpose which influence reading behavior.

Matching items in testing reading comprehension can be useful for testing vocabulary in context. According to Brown (2004) matching words are used for the beginner level testing in which students match together items appropriately (e.g. antonyms, synonyms and interpreting different signs or labels). Like other techniques matching items have their own pros and cons. Brown (2004) sees matching items biggest advantage as they offer an alternate to an MCQ or fill in the blanks items. Matching items are relatively easy to construct if they are

constructed carefully but learners in their real life rarely encounter such activity.

Picture-Cued tests are also used at beginner level with two variations. Learners are given a passage and they choose the described pictures (Brown, 2004). Kitao and Kitao (1997) added that this activity may be altered to challenge higher cognition level in such a way that the student looks at a picture or a number of pictures and decides the sentences describe the picture accurately. IELTS reading test also uses diagram labeling as a way to test reading ability. There is another way to test reading by a cloze test and open-ended responses. Kintsch and Yarbrough (1982) investigated the effects of cloze test and open-ended response on reading comprehension. They find that open-ended response measures only test takers' comprehension of main ideas of the reading text while cloze test touches upon students' local understanding (local meanings) and they are unable to reflect over all comprehension of the text. Pishghadam and Tabataba'ian (2011) findings show that exhausting the task linguistically by means of textual organization, additional clues with a number of gaps results in better performance. Moreover, this is a classical method and is regarded as best to test reading ability.

It has been long argued that a cloze test measures textual knowledge which involves the cohesion of a text i.e. global understanding and this test can be applied to both elementary and advance levels. According to Heaton (1990) summary writing deals with whole-text, super macro-level skill. Brown (2004), Heaton (1990) and Weir (1990) observed advantages and disadvantages in this testing technique i.e. a cloze test serves as a diagnostic tool and several researchers have applied this tool both in cognitive processing of sentences and in practical language teaching (Al-Shumaimeri, 2006). This type of test can be used in loose structure of sentences, in which local information is required, or in which global information is required i.e. macro-structural level. Greene (2001) reports a cloze test as a valid measure of reading comprehension in order to check macro level comprehension and also global coherence. Shanahan, Kamil and Tobin (1982) compared different cloze tests' natural, intruded and scrambled and concluded that cloze test could not be used for global comprehension. Mckenna and Layton (1990) have observed that a cloze test measures abilities at an inter-sentential level like as TOFEL uses sentence insertion, and summary completion. Hicks and Monroe (1979) explicate through their findings that there is no strong relationship between a cloze test and informal reading texts. Weir (1990) states that in open-ended items right answers are never provided as compared to MCQ tests and guessing is almost eliminated. Brown (2004) finds that open-ended questions enhance creativity among students by using their expressions. This may start follow-up discussions and have positive wash back effect. The main disadvantage of using this item is the teacher who may evaluate their writing skills rather than reading ability. Nuttall (2005) mentions that the teacher asks all those questions with possible foreseeable answers otherwise the student may confuse to search answers. Moreover, scoring is very much subjective. Teacher faces difficulty to mark correct, partially correct and wrong answers. There is no one particular testing item format which measures all skills and sub skills simultaneously. Test of reading combines different item formats to check both high and low order skills. Multiple choice items and cloze item type is majorly used in testing reading comprehension as well as by high stake testing bodies i.e. IELTS, GRE and TOFEL.

1.5 Background Knowledge

Researches support the significance of learners' background knowledge about a topic as it increases reading comprehension and memory. Kintsch (1988) used a theory i.e. constructive-integration process which describes an important role in text comprehension through background knowledge. He states that when background knowledge combines with informational text, it constructs a situational model. The situational model constructs fully when the reader has more background knowledge of the text. Eskey (1986) mentioned two categories of the schema i.e. form and substance. Knowledge of form provides linguistic information about semantic and syntactic patterns of the text, whereas, knowledge of substance provides information about the text with respect to pragmatic, cultural and information about a particular topic. Brandão and Oakhill (2005) stated that the reader must use prior knowledge to understand the implicit information of a topic to get better results. Studies, related to reading comprehension and background knowledge, have been conducted. The topics were chosen from specific fields and cultures in which positive wash back was reported by the researchers. Instead of topics selected from outside, their cultural and personal life settings resulted in negative wash back. The selected topics were experienced by the students in their real life situations (Aleptekin, 2006; Chen & Donin, 1997; Lee, 2007; Yuet & Chan, 2003). Previous studies have also showed that learners' text comprehension is enhanced through high

background knowledge (Taft & Leslie, 1985). Pearson, Hansen and Gordon (1979), Marr and Gormley (1982) have observed that struggling readers comprehend texts often on a topic through their high level of background knowledge. Readers' having high background information constructs detailed situational model and comprehend more precisely. The two main factors in constructing this model are background knowledge of reader and demands of the reading task (Blanc & Taperio, 2001).

The main aim of reading comprehension is to integrate background knowledge and textual information. There may be poor comprehension that leads to weaker development of situational model in case of overusing of two factors i.e. background and textual information (Brandão & Oakhill, 2005; Canin & Oakhill, 2001; Kamalski, Sanders & Lentz, 2008; Kintsch, 1988; Pudilo, 2007). Brandão and Oakhill (2005) conducted a research to investigate the impact of background knowledge on narrative texts comprehension. Young children were the participants of the study and were asked how they did get the answers of different comprehension questions. The results showed that they got their answers from the text and also overused their background knowledge. In the findings of the study, only 6.46% students used their background knowledge to answer the questions rather than textual information. The results determined that those students were not constructing situational model as described in Kintsch (1988). They were also not bridging new information with old information. Pudilo (2007) conducted the study similar to that of Brandão and Oakhill (2005), he concluded that the development of mental representation of a text in a coherent way involves interaction between background knowledge and explicit textual information.

Interaction between textual information and prior knowledge has a high impact on the reading ability of a reader to make inferences. Several other researchers found that high level of background knowledge about a topic indicated fourth grade student's ability to make inferences instead of reading comprehension skills (Marr & Gormley, 1982). In some cases, it is not a matter of background knowledge that has an impact on inference and comprehension. Rather, it is the ability to integrate old information with new information given in the text. Lin (2002) investigated the point of view of EFL learners on background knowledge and also its impact on comprehension. The findings of the study showed that majority EFL learners considered background knowledge to reading comprehension. According to middle school learner's reader's prior knowledge is the most type of background while reading English text. In terms of Eskey (1986) this linguistic knowledge is the knowledge of form. On the other hand, tertiary students state that socio-cultural knowledge is the most type of prior knowledge (Lin, 2002). In line to Lin (2002), Garth-McCullough (2008) researched background knowledge of African American students to investigate the relationship between culture and reading comprehension. The results showed that the students with high background knowledge of their own culture can understand others' culture in a better way while reading (Garth-McCullough, 2008). A number of studies associated with background knowledge and reading comprehension reported positive wash back for specific disciplines (Chen & Donin, 1997) and cultures (Lee, 2007; Yuet & Chan, 2003). In these studies, topics were selected from specialized fields and particular cultures and students experienced real life situations with these topics. While few studies took topics from outside of particular discipline and cultures. Another important impact of background knowledge on comprehension is the problem of test bias. Thus, it can be said that it is nearly impossible to create test without biasness which is created by background knowledge. It would seem unwise to do, as background knowledge is an essential component of reading comprehension because there is no such measuring tool as to see what resides in brain. Therefore, reading assessment is always inference based.

2. Conclusion

In this study, various principles on the construction of reading tests have been explored in the light of authentic researches and practices adopted by high stake testing bodies like IELTS, GRE, TOFEL. Findings showed that text selection, text types, test items, and background knowledge in testing reading abilities are different according to the cognition level of learners. Text selection is one of the main features for testing reading comprehension. Literature significantly discussed that text selection should be made according to the test takers' learning objectives, close to real life, interest, level, textual organization and their cultural schema. Selection of a text with great linguistic complexities with respect to cultural appropriateness and objectives then it is possible that due to linguistic barriers the learners might not be able to perform better. Subsequently, it is better to select a text and categorize it according to the low and high order thinking skills. At times, one text is used for testing different levels by varying the degree of complexity. The text type and text selection principles devised by Brown

(2004) and international testing bodies like IELTS, TOFEL and GRE are taking it helpful for test constructors. These suggested text types are helpful for learners to extract appropriate meanings by using their background knowledge (Brown, 2004). Simple narrative passages are useful for elementary level then the same text can be made complex by using more vocabulary, connected phrases and relative clauses, and metaphors. Hence, the text type at elementary level can be descriptive and narrative i.e., newspaper, greetings, and magazines but at advance level, expository and argumentative text types can be chosen like professional writings, lab reports, and technical writings according to their socio-cultural interests. Text length is varied at different levels. At elementary level it can be 250-300 words, at intermediate level 300-800 and at advance it can be 1200 words or more. Text length includes shorter or longer texts. If the test constructor, constructs test of text length of 1000 words then it is feasible for the test constructor to incorporate a number of items and students with low proficiency level can perform better. Contextual clues and discourse markers facilitate reading comprehension. If shorter texts are given then due to limited number of clues show less proficiency.

The test formats, used in testing reading, are MCQ, cloze test, open-ended response, filling gaps, and true false. The most debateable formats are true/false and cloze test with multiple choices for testing reading in IELTS and TOFEL. If we are using true/false items for testing reading comprehension, we may check first two levels of cognition of Bloom's taxonomy but in real life they are not useful. Researchers showed highest difference in using these formats, and test takers performed better at MCQ items. While, some researchers are in favour of cloze tests but some strongly recommend MCQ items. If we take cloze test and apply framework of Bloom's taxonomy, it is seen that cloze test works up to fourth level of analysis. But MCQ tests work till evaluation level in which the test taker infers implied meanings of author. Researchers who favour the cloze test, are testing reading with summary cloze and close elide rather than standard cloze. In this way, MCQ tests are more valid rather than cloze ones. Nowadays, mixed test formats are used by testing organizations. If we are using one item format to test all the abilities of the test takers than it is impossible to test with one particular format. Testing reading is an integrated test of abilities. Therefore, text selection and items format should be decided according to the purpose of testing.

3. Pedagogical Implications

Following pedagogical implications are proposed for ELT practitioners to test reading abilities:

1. The text selection should be different for different levels. While selection, different genres serve different purposes of the tests. Narrative writings like short stories, short passages are suitable for elementary level. At intermediate level descriptive writings like technical writings, map reading, letters, emails, and historical events' description can be given. At advance level expository and argumentative essays, reports, immigration documents, instructional manuals can be selected
2. The text length should comprise of 250-300, 400-850, and more than 1200 words at elementary, intermediate, and advanced levels respectively.
3. For beginners true false, matching, picture cued sentences' test items should be given to test knowledge and comprehension level. At intermediate level summary cloze, cloze elide and information transfer activity can be given. At advance level, multiple choice items should be given to test their higher cognition level and those type of items should be included which involve authors' implied meanings and test takers have to infer meanings from it.

References

1. Alderson, J. C. (1990). Testing reading comprehension skills (Part One). *Reading in a Foreign*
a. *Language*, *6*(2), 425-438.
2. Al-Shumaimeri, Y. (2006). The Effects of Content Familiarity and Language Ability on Reading
a. Comprehension Performance of Low and High-ability Saudi Tertiary Students Studying English as a Foreign Language. *Education Science and Islamic Studies, 18*, 1-19
3. Alptekin, C. (2006). Cultural familiarity in inferential and literal comprehension in L2 reading.
a. System, 34, 494-508.
4. Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
5. Alderson, J. C., & Lukmani, Y. (1989). Cognition and reading: Cognitive levels as embodied in
a. test questions. *Reading in a Foreign Language*, *5*(2), 253-270.
6. Andreassen, R., & Bråten, I. (2010). Examining the prediction of reading comprehension on
a. different multiple- choice tests. *Journal of Research in Reading*, *33*(3), 263-283.

7. Arias, I. J. (2007). Selecting reading materials wisely. *Letras*, *41*, 131-151.
8. Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford
a. University Press.
9. Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford
a. University Press.
10. Berkoff, N. A. (1979). Reading skills in extended discourse in English as a foreign language.
a. *Journal of Research in Reading*, *2*(2), 95-107.
11. Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy*
a. *of educational objectives: The classification of educational goals: Handbook I: Cognitive domain* (No. 373.19 C734t). New York, US: D. Mckay.
12. Brandão, A. C. P., & Oakhill, J. (2005). "How do you know this answer?"–Children's use of text
a. data and general knowledge in story comprehension. *Reading and Writing*, *18*(7-9), 687-713.
13. Brown, D. H. (2004). *Language assessment, principles children classroom practices.* New York:
a. Longman.
14. Blanc, N., & Tapiero, I. (2001). Updating spatial situational models: Effects of prior knowledge
a. and task demands. *Discourse Processes*, *31*(3), 241-262.
15. Carter, M., Rastatter, M. P., Walker, M. M., & O'Brien, K. (2009). The effects of frequency
a. altered feedback on the reading processes of adults with reading disorders. *Neuroscience Letters*, *461*(2), 69-73.
16. Carver, R. P. (1990). *Reading rate: A review of research and theory*. New York: Academic
a. Press.
17. Chan, N., & Kennedy, P. E. (2002). Are multiple-choice exams easier for economics students? A
a. comparison of multiple-choice and "equivalent" constructed-response exam questions. *Southern Economic Journal*, *68*(4), 957-971.
18. Chen, Q., & Donin, J. (1997). Discourse processing of first and second language biology texts:
a. Effects of language proficiency and domain-specific knowledge. *The Modern Language Journal*, *81*(2), 209-227.
19. Chen, C. (2010). On reading test and its validity. *Asian Social Science*, *6*(12), 192-195.
20. Cohen, A. D. (2007). Test-taking strategies. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. E.
a. Turner & C. Doe (Eds.), *Language Testing Reconsidered* (pp. 89-111). Ottawa: University of Ottawa Press.
21. Cutting, L. E. (2017). *Do all reading comprehension tests assess equally?* Retrieved April 11,
a. 2020 from https://www.hmhco.com/blog/do-all-reading-comprehension-tests-assess-equally.
22. Cain, K., Oakhill, J.V., Barnes, M.A., & Bryant, P.E. (2001). Comprehensions kill, inference-
a. making ability, and their relation to knowledge. *Memory and Cognition*, *29*, 850-859.
23. Davis, F. B. (1968). Research in comprehension in reading. *Reading Research Quarterly*, *3*(4),
a. 499-545.
24. Day, R. R. (1994). Selecting a passage for the EFL reading class. *English Teaching Forum*,
a. *32*(1), 20-23.
25. Day, R. R., & Park, J. S. (2005). Developing reading comprehension questions. *Reading in a*
a. *Foreign Language*, *17*(1), 60-73.
26. Downing, S. M. (2006). Selected-response item formats in test development. In S. M. Downing
a. T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 287-301). Mahwah: Lawrence Earlbaum Associates, Publishers.
27. Engelhard, J. G. (2001). Historical view of the influences of measurement and reading theories
a. on the assessment of reading. *Journal of Applied Measurement*, *2*(1), 1-26.
28. Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B.,
a. &Brosvic, G. M. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record*, *52*(2), 187-201.
29. Erten, İ. H., & Razi, S. (2009). The effects of cultural familiarity on reading comprehension.
a. *Reading in a Foreign Language*, *21*(1), 60-77.
30. Eskey, D. E. (1986). Theoretical foundations. In F. Dublin, D. E. Eskey & W. Grabe (Eds.),
a. *Teaching Second Language Reading for Academic Purposes* (pp. 3-23). Reading: Addison-Wesley.
31. Garth-McCullough, R. (2008). Untapped cultural support: The influence of culturally bound
a. prior knowledge on comprehension performance. *Reading Horizons: A Journal of Literacy and Language Arts*, *49*(1), 1-30.

32. Gebhard, J. G. (2006). *Teaching English as a foreign or second language: A teacher self-*
a. *development and methodology guide*. Ann Arbor, US: University of Michigan Press.
33. Grabe, W. (1991). Current developments in second language reading research. *TESOL*
a. *Quarterly*, *25*(3), 375-406.
34. Grabe, W., & Stoller, F. (2002). *Teaching and researching reading.* New York: Longman.
35. Güler, H. (2007). *Non-native EFL teachers' beliefs about teaching reading* (Master's thesis).
a. Anadolu Üniversitesi, Eskişehir, Turkey.
36. Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. London:
a. Routledge.
37. Heatherington, A. (1985). Assessing the suitability of reading materials for ESL students. *TESL*
a. *Canada Journal, 3*(1), 37-52.
38. Heaton, J. B. (1975). *Writing English language tests: A practical guide for teachers of English*
a. *as a second or foreign language*. London: Longman.
39. Heaton, J. B. (1988). *Writing English language tests*. London: Longman.
40. Heaton, J. B. (1990). *Classroom testing*. London: Longman.
41. Hedgcock, J. S., & Ferris, D. R. (2018). *Teaching readers of English: Students, texts, and*
a. *contexts*. London: Routledge.
42. Hicks, R. D., & Monroe, E. (1979). *A comparison of reading achievement, current reading*
a. *placement, sex, age, intelligence, informal reading inventory, and the cloze procedure.* ERIC Document Reproduction Service No. Ed. 181438.
43. Hudson. T. (2007). *Teaching second language reading*. Oxford: Oxford University Press Oxford.
44. Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University
a. Press.
45. Kamalski, J., Sanders, T., & Lentz, L. (2008). Coherence marking, prior knowledge, and
a. comprehension of informative and persuasive texts: Sorting things out. *Discourse Processes*, *45*(4-5), 323-345.
46. Kim, A. Y. (2009). Investigating second language reading components: Reading for different
a. types of meaning. *Columbia University Working Papers in TESOL & Applied Linguistics, 9*(2), 1-28.
47. Kintsch, W., & Yarbrough, J. C. (1982). Role of rhetorical structure in text comprehension.
a. *Journal of Educational Psychology*, *74*(6), 828-834.
48. Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-
a. integration model. *Psychological Review*, *95*(2), 163-182.
49. Kitao, K., & Kitao, S. K. (1997). Selecting and developing teaching/learning materials. *The*
a. *Internet TESL Journal*, *4*(4), 20-45.
50. Kobayashi, M. (2002). Method effects on reading comprehension test performance:
a. Textorganization and response format. *Language Testing, 19*(2), 193-220.
51. Kobayashi, M. (2005). An investigation of method effects on reading comprehension test
a. performance. A paper presented at the *Proceedings of the 4tAnnual JALT Pan-SIG Conference*, Tokyo, Japan.
52. Lee, S. K. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students'
a. reading comprehension and learning of passive form. *Language Learning*, *57*(1), 87-118.
53. Lin, Z. (2002). Discovering EFL learners' perception of prior knowledge and its roles in reading
a. comprehension. *Journal of Research in Reading*, *25*(2), 172-190.
54. Linn, R. L., & Miller, M. D. (2008). *Measurement and assessment in teaching*. New Delhi,
a. India: Pearson Education.
55. Liu, F. (2011). A short analysis of the text variables affecting reading and testing reading.
a. *Studies in Literature and Language*, *2*(2), 44-49.
56. Lumley, T. (1993). The notion of sub skills in reading comprehension tests: An EAP example.
a. *Language Testing*, *10*(3), 211-234.
57. Marr, M. B., & Gormley, K. (1982). Children's recall of familiar and unfamiliar text. *Reading*
a. *Research Quarterly*, *18*(1), 89-104.
58. Matthews, M. (1990). Skill taxonomies and problems for the testing of reading. *Reading in a*
a. *Foreign Language*, *7*(1), 511-517.
59. McKenna, M. C., & Stahl, K. (2015). *Assessment for reading instruction*. New York: Guilford
a. Press.
60. McNamara, T., & Roever, C. (2006). *Language testing: The social dimension.* Malden:

a. Blackwell.
61. McKenna, M. C., & Layton, K. (1990). Concurrent validity of cloze as a measure of
a. intersentential comprehension. *Journal of Educational Psychology, 82*(2), 372–377.
62. Munby, J. (1978). *Communicative syllabus design.* Cambridge: Cambridge University Press.
63. Marzano, R., & Kendall, J. (2007). *The new taxonomy of educational objectives* (2nd ed.).
a. Thousand Oaks, CA: Corwin Press.
64. Nevo, N. (1989). Test-taking strategies on a multiple-choice test of reading comprehension.
a. *Language Testing*, *6*(2), 199-215.
65. Nuttall, C. (1996). *Teaching reading skills in a foreign language.* Oxford: Heinemann
a. International.
66. Okumura, N. (1998). The variances in L2 reading ability in accordance with text type:
a. Expository and narrative materials. *Surcle, 1*, 16-25.
67. Osterlind, S.J. (1998). *Constructing test items: Multiple-choice, constructed-response,*
a. *performance and other formats* (2nd Ed.). New York: Kluwer Academic Publishers.
68. Paxton, M. (2000). A linguistic perspective on multiple choice questioning. *Assessment &*
a. *Evaluation in Higher Education*, *25*(2), 109-119.
69. Pearson, P. D., Hansen, J., & Gordon, C. (1979). The effect of background knowledge on young
a. children's comprehension of explicit and implicit information. *Journal of Reading Behavior, 11*(3), 201-209.
70. Pearson, P. D., & Johnson, D. D. (1978). *Teaching reading comprehension.* New York: Holt,
a. Rinehart and Winston.
71. Pishghadam, R., & Tabataba'ian, M. (2011). IQ and test format: A study into test fairness.
a. *Iranian Journal of Language Testing, 1*(1), 17-29.
72. Pulido, D. (2007). The relationship between text comprehension and second language incidental
a. vocabulary acquisition: A matter of topic familiarity? *Language Learning, 57*, 155-199.
73. Rezaei Ghahroudi, M., & Sheikhzadeh, E. (2017). Selecting reading texts for university Iranian
a. EFL students. *Journal of Advances in English Language Teaching, 5*(3), 25-30.
74. Rost, D. H. (1993). Assessing different components of reading comprehension: Fact or fiction?.
a. *Language Testing*, *10*(1), 79-92.
75. Sainsbury, M., Harrison, C., & Watts, A. (2006). *Assessing reading: From theories to*
a. *classrooms.* Slough: National Foundation for Educational Research.
76. Scheuneman, J. D., & Gerritz, K. (1990). Using differential item functioning procedures to
a. explore sources of item difficulty and group performance characteristics. *Journal of Educational Measurement*, *27*(2), 109-131.
77. Shanahan, T., Kamil, M. L., & Tobin, A. W. (1982). Cloze as a measure of intersentential
a. comprehension. *Reading Research Quarterly*, *17*(2), 229-255.
78. Shohamy, E. (1984). Does the testing method make a difference? The case of reading
a. comprehension. *Language Testing, 147*(1), 147–170.
79. Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new
a. research and practice paradigms in the testing of English-language learners. *Educational Researcher*, *32*(2), 3-13.
80. Taft, M. L., & Leslie, L. (1985). The effects of prior knowledge and oral reading accuracy on
a. miscues and comprehension. *Journal of Reading Behavior*, *17*(2), 163-179.
81. Urquhart, A. (1998*). Reading in second language: Process, product and practice.* New York:
a. Longman.
82. Vacca, R. T., & Vacca, J. A. L. (2005). *Content area reading: Literacy and learning across the*
a. *curriculum, 8/e.* Boston: Allyn & Bacon.
83. Wallace, C. (1992). *Reading.* New York: Oxford University Press.
84. Walter, C. (1992) *Authentic reading.* Cambridge: Cambridge University Press.
85. Weir, C. J. (1990). *Communicative language testing*. London: Prentice Hall.
86. Weir, C. J. (1993). *Understanding and developing language tests*. New York: Prentice Hall.
87. Weir, C. J., & Porter, D. (1994). The multi-divisible or unitary nature of reading: The language
a. tester between Scylla and Charybdis. *Reading in a Foreign Language*, *10*(2), 1-19.
88. Wolfe, M. B., & Woodwyk, J. M. (2010). Processing and memory of information presented in
a. narrative or expository texts. *British Journal of Educational Psychology*, *80*(3), 341-362.

89. Yuet, C., & Chan, H. (2003). Cultural content and reading proficiency: A comparison of
a. mainland Chinese and Hong Kong learners of English. *Language Culture and Curriculum*, *16*(1), 60-69.
90. Zheng, Y., Cheng, L., & Klinger, D. A. (2007). Do test formats in reading comprehension affect
a. second-language students' test performance differently? *TESL Canada Journal*, 65-80.