International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 07, 2020
ISSN: 1475-7192

The Process of Developing a Reading Test: A Review Article

[1]Kalsoom Jahan, [2]Sharjeel Ashraf, [3]Sumaira Mukhtar

Abstract

This review article discusses the criteria for developing tests for reading skills. Reading skills are input language skills whichare considered as the most basic skills in language learning.The text focuses on the basic approaches used in developing reading tests, selection criteria of texts and authenticity, selection of test items, and the role of Bloom taxonomy in testing reading skills. Asian journal of language testing, Language testing international,Journal of research in reading,Reading research quarterly, Reading for professional purposes, Journal of reading, Journal of reading behaviour, and Language testing are reviewed to answer the research questions.Discussions are analysed and concluded that the level of the learner, objectives, and text itself decide the types of questions that can be constructed and included in reading tests.Different text may tend to different test items such as true-false, multiple-choice, matching, provide information in the table, and yet to add open-ended questions.

Keywords: *Bloom taxonomy, Cognitive process, Reading skills,Reading comprehension, Testing*

I. Introduction

Reading skills are closely connected with other language skills as the ability to recognize words, associating sounds with their equivalent graphic symbols. Reading comprehension is defined as "understanding a written text means extracting the required information from it as efficiently as possible" (Grellet, 1981, p.3).According to J.BHeaton (1990), reading is linked with the inferring the meaning of the word by understanding word-formation processes like derivation, roots affixation, and compounding. Contextual evidence is also important in this regard as one of the members of the group reveals the plot and understanding the links within the sentence, particularly the features of sentence structure, fronting, and theme, negation, embedding, and complex. To comprehend the relation between the parts of the text through lexical devices as synonyms, repetition, and grammatical cohesive devices especially cataphoric and anaphoric references, antithesis, and connectives. Reading ability also includes the perception of the spatial and temporal relationship, and also the sequencing of concepts.

In reading testing, a hot discussion deals with the potential sources of local test items rely on common items types, the use of different items to measure the same sub-skills of reading, and a single passage is used

[1] Lecturer in English, Lahore Garrison University, Lahore, Pakistan
[2] Lecturer in English, Lahore Garrison University, Lahore, Pakistan
[3] Lecturer in English, Lahore Garrison University, Lahore, Pakistan

followed by a set of items. Andrich, Humphry, and Marais (2012) proposed a multidimensionalconstruct, the use of multiple items types within a test like open-ended and close-ended items in a test. Kobayashi, (2002), and Shohamy(1984) has suggested that numerous item types measure different skills of reading comprehension.

Figure 1 *The following figure is adapted from Longman Handbooks for Language Teachers: Writing English Language Tests (J.B Heaton, 1990, p. 106).*

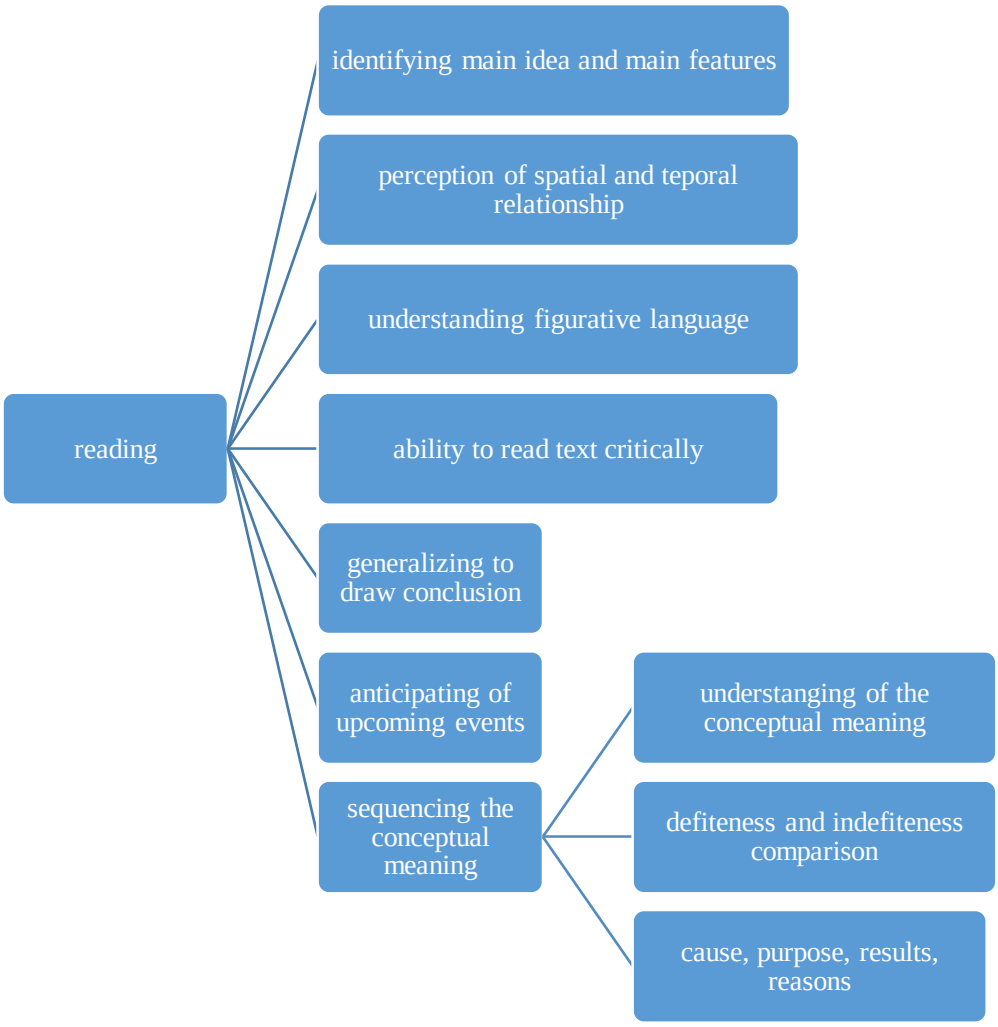

Figure 1 shows reading skills concerning its sub-skills which are involved in reading comprehension. All these skills are important in terms of developing a reading test. There are different factors like level, interest, and needs of the students, aims, and objectives of the tests which have to be kept in mind while constructing a reading test.

Research questions

1. Whichitem sequencing approaches have been used in testing reading skills?
2. What have been the criteria for text selection in testing reading skills?
3. What have been the waysof constructing items in reading tests?
4. What have been the aspects of text and authenticity in testing reading skills?
5. What has been a multidimensional construct used in reading testing?

6. What factors have been affecting reading testing?
7. What has been the effect of the cognitive process in reading tests?
8. What levels of bloom taxonomy have been focused on testing reading skills?

II. Methodology

Data was collected from different sources to answer the research question. The Journal of language testing was used for recent researches. *Asian Journal of Language Testing, Language Testing International,Journal of Research in Reading, Reading Research Quarterly,Reading for professional Purposes, Journal ofReading,Journal of Reading Behaviour,*and *Language Testing* are reviewed. These international journals publish original research and review articles on language testing and assessment. It provides a forum for the exchange of ideas and information between people working in the fields of first and second language testing and assessment. This includes researchers and practitioners in EFL and ESL testing and assessment in child language acquisition and language pathology. Google Scholars, Sage journals, Willy online, and J. Store were also used to find out more relevant and detailed material. Different keywords were used in searching related research articles such as what is reading testing, basic criteria for reading testing approaches for testing reading, selections of items in a reading test, and role of Bloom Taxonomy in reading assessment, etc. Reading testing journals were given main preference, secondary preference was given to the journals of language testing and then journals of ELT and then the journals of education and so on. During compiling and retrieving research files research ethics were kept in consideration.

III. Finding and Discussion

This section is dealing with evidence that supports, denies, and rebuttal various notions, approaches, and theories regarding reading skills and reading testing. Results are generated after reviewing research articles, thesis, research reports, and books. Research ethics were kept into consideration while reviewing and citing this research.

3.1 Approaches Used in Reading Testing

The use of scheme theory and background knowledge in reading activities are very important in promoting the reading comprehension of the students. Different approaches are used in reading comprehension and testing reading skills.

3.1.1 Bottom-up treatment. In bottom-up treatment, the comprehension starts with the very basic component of languages such as words understanding word building or word formation and grouping of words. Here the main focus was given on textual decoding,

3.1.2 Top-down treatment. In Top-down treatment, the main emphasis was given on reader understanding of the text and prior knowledge and then moving to the discrete points of the text.

3.1.2 Interactive treatment. The interactive treatment approach involves both textual codes as mentioned in the bottom-up approach and background knowledge is working interactively.

(Chou, 2011;Grabe and Stoller, 2002; Pulido and Hambrick, 2008;).

So, in constructing a reading test, the approach can be selected according to the level of the learner, interest of the learner, and the type of the test of reading.

3.2 Text Selection in Reading Tests

The selection of text in reading tests involves multiple aspects such as length of the text, type of the text, level, and interest of the students.According to J.B Heaton (1990) in reading tests, it is useful to include different types of texts for reading comprehension including literary prose extracts: e.g. newspaper articles, instruction manuals for the usage ofmachinery and appliances, directory extracts, timetables, public notices, and maps advertisements, etc. By including these types of text will not only deliver more reliable and realistic means of assessment and evaluation but will also help to encourage students by representing how to target language which is used in a real-life situation. Alderson (2000) also supports the idea of including literary text in reading tests such as drama, prose, short stories, and poetry.

The main point to consider inconstructing a test of reading comprehension is to select aimsthat a test is going to measure. Spolsky (1985) finds that testing items depend upon the things which are going to be measured.Hughes (1989) discusses for both macro-skills like scanning a text to findspecific information and micro-skills are those that are using context to predict the meaning of unfamiliar words, as having significance for the assessment of reading skill. He concludes that the skills that demonstrate command of the macro-skills also suggest mastery of the micro-skills. The researcher was fascinated by developing a test to assess how well learners would be able to accessmeaning from an authentic text by using the platform of the test questionsthat was the reason for the researcher to think to use an integrative approach in testing reading comprehension skills.

The selected text must include such information that can be scanned by the use of different test items. Hughes (1989) suggests that the selection of text for reading test includes passages which contain enough discrete information for scanning if needed. He also advises that a fresh start must be given to the student by providing him several passages. According to him, three to four passages must be included a reading test with multiple test items/questions.Jafapur (1987) suggests that the text must provide the opportunity to use a short context technique. He claims it will help in measuring reading skills rather than anything else. It will also point out the relevant real behavior of the world. The authenticity of the text is a major aspect of the text in text selection in developing a reading test.

3.3 Aspects of Text and Task Authenticity

The authenticity of the text involves students in life-like situations. Hill and Parry (1994) mention that the text used in reading tests are not authentic all the time. There is a degree of resemblance between the original text and the passage used in a reading comprehension text, it raised questions upon the authenticity of the text used in the context of the test. It also highlights the fact that several authentic texts that are used in reading comprehension tests are not always in facsimile form, it also denies cues such as format and typeface. Although Campbell (2005) suggests the text used in the reading comprehension test must form amount to three pages in its final form. He felt that this is a justified form of the text and seems to be as the original, and this is the task and students may have to tackle it in the actualreal world. Grabe and Stoller(2002) purport the idea that test can be divided into sections to make reading more comprehensive, and the students must be provided with one page at a time to find out the answers, the questions must be accompanied with the directions

which can be helpful for the learner to scan information that is required by the examiner. Student can be facilitated in this way, after reading a small amount of text, they would be able to answer the questions.The language is also important in designing authentic tasks in reading comprehension tests. Hughes (1989) suggests that the questions included in reading comprehension tests must always be less demanding as compared to the text itself. Hill and Parry (1994) support the idea that if it is possible to write questions in the test takers' language, so it should be in the test taker's language. Kobayashi (2002) supports by saying that it can make the test lifelike and more practicable for the student and the teacher to stay with the test and not giving up.

3.4 The Ways of Constructing Items in Reading Tests

Constructing test items is a crucial task in reading testing and different researches have purports different suggestions. Heaton (1990) suggests open-ended questions and multiple-choice items are more suitable for the reading comprehension tests. Moreover, he prefers multiple-choice items in testing reading skills. In many ways the text itself is a basic point that helps in determining the types of questions, different text may lend different items and such as a text may lend to true-false items, others to multiple-choice items, others to matching items, others to others to provide information in the table and yet to add open-ended questions. Most of the time, two or more than two items can be constructed for a text. Many researches examine the influence of item format and the range of success variation related to changes in testing context and methodology. Some of these studies favoured that a large part may be equally targeted with a constructed response and multiple-choice items. There are different opinions as well, such as Campbell (2005) asserted that in literary text multiple-choice items should be avoided. Haladyna and Rodriguez (2013) prefer different items in a reading comprehension testbecause different items have different cognitive functions in the test. Jahan, Shah and Ahmad (2019) favoured to add communicative factors in testing reading skills to perform these cognitive functions. So, these testing items should include both communicative and cognitive functions.

It is difficult to give a preference for one testing item to the other.For instance, Van den Broek (2012) mentions that it is impossible to differentiate the use of open-ended questions or multiple-choice items to measure the substantial differences in intellectual abilities. Farr, Pritchard, and Smitten (1990) add by saying that although the approaches used in answering multiple-choice items may matchless to non-test situation reading, these types of items have much resemblance to the common reading in school, in which texts are often skimmed and scanned to find out particular information. However, in testing a particular skill, the item process validity should finally be considered. Pearson and Hamm (2005) conclude that using of multiple items emerge when tasks call for a deeper cognitive engagement of the students, such as expecting students to consider numerous stances towards a text (multiplicity), or to create a link among ideas across different types of texts (intertextuality).

Gap-filling techniques and cloze tests are other alternatives to standard multiple-choice items. In gap-filling items, some words are omitted from the text and it is expected from the students to fill in the gap with the suitable word from a list of alternatives. Kintsch and Yarbrough (1982) mention that it has been questioned that how much these skills are reliable in testing higher-order skills or these items are restricted in measuring skills at micro-level only. The cognitive ability of the learner is dependent on the gaps provided in the text only, soAlderson (2000) purports the idea that it is not easy to measure higher-order skills as global comprehension, syntactic knowledge and lexical access of the text in the test by using only cloze test items and

gap-filling items, it requires multiple items in the test. Yamashita (2003) has asserted thatstill, thesetypes of test items have been proven to differentiate well between skilled and less skilled readers.

3.5 A Multidimensional Construct in Reading Testing

To assess and evaluate reading comprehension, it has been a critical component in all forms of educational quality, though it's a matter of national testing or internationally organized one. Van den Broek (2012) asserted that reading comprehension has been a challenge and multidimensional construct are favored most. Andrich, Humphry, and Marais (2012) proposed a multidimensional construct, the use of multiple item types within a test like open-ended and close-ended items in a single test. Kobayashi, (2002), and Shohamy(1984) has suggested that numerous item types measure different skills of reading comprehension.

Leslie and Caldwell (2009) There are different contextual factors such as type of the text, its topic and purpose, yet these have to be managed in balancing different subskills of reading such as inferring of explicit information, critical evaluation, and interpreting content. It affects the comprehension measures, variances, and applicability of test items.

3.6 Factors Affecting Reading Testing

Many studies have been devoted to recognizing the factors that best explain the difference incomprehension. Cutting and Scarborough (2006) mentions that in exploring the subdivisions of the construct and inter-correlations among items various factors have been identified. According to many types of research, different test items are used to measure different skills at different cognitive levels in the reading comprehension test. Nation and Snowling (1997) stated that the length of the text and the number of passages of text including in test effects the performances of the students. Cutting and Scarborough (2006) have noted that these differences clearly affect the range of possible inferences about the proficiency level of the students and especially when it encounters the reading difficulties. He argued that reading comprehension assessment and evaluation would be different in the future, such as comparative studies that can be conducted on existing frameworks for assessment including evaluating items properties, usability, dimensionality, and generalizability. Stress would be given to the deep analysis of the test items of every component of reading in terms of processing, and cognitive strategies which have been considered as existing assessment devices.

3.7 Effect of Cognitive Process in Reading Tests

The cognitive level of the students involves in different test items, as Van den Broek (2012) commented that in reading comprehension cognitive processes involved as a particular component. Although, the practicalities of a specific test are well informed in terms of passage length, time allotment, item format, and reliability measures. Though less information has been provided about the cognitive process involved in different test items. It also reflects the theoretical framework of reading comprehension that is highly influenced by Bloom's (1956) taxonomies or the hierarchical framework of cognitive domains of knowledge. Alderson (2000) Pearson and Hamm, (2005) and Weir (2013) have preferred the use of the cognitive framework of bloom in constructing different test items according to the different cognitive levels to assess the complexity in pieces of information, inferring meaning and evaluation of text critically. Thus, a categorization by cognitive targets can be helpful to ensure cognitive domain coverage. Test scores would not be uncertain if reliably reported by subscales of cognitive process levels.

3.8 Reading Assessment Tests and Levels of Bloom Taxonomy

Bloom (1956)recommended six different cognitive stages in learning from simple recall, knowledge, or recognition of facts, is considered as the lowest level, and the highest order, which is evaluation. Bloom taxonomy helps in developing questions in reading tests according to the cognitive levels of learning.

Table: 1 Levels of Bloom's Taxonomy vs selection of test items.

Level	Title	Key Words	Types Of Questions
1.	Knowledge	who, what, why,	What is . . . ?
	(Recalling facts,	when, which,	Where is . . . ?
	terms, basic	choose, find, how,	When did _____ happen?
	concepts and	omit, where define,	When did . . . ?
	Answers.)	label, show, spell,	Can you recall . . . ?
		list, match, name,	How did _____ happen?
		relate, tell, recall,	Why did . . . ?
		Select	How would you show . . . ?
			How would you explain . . . ?
			Which one . . . ?
			How would you describe . . . ?
			Who was . . . ?
			Who were the main . . . ?
			Can you select . . . ?
			Can you list three . . . ?
2.	Comprehension	compare, contrast,	How would you classify the type of .. ?
	(Demonstrating	demonstrate, extend,	How would you compare . . . ?
	understanding	illustrate, infer,	Which statements support . . . ?
	of facts and	outline, interpret,	Can you explain what is happening . . .

	ideas by organizing, comparing, translating, interpreting,giving Descriptions and stating main ideas.)	explain, relate, rephrase, translate, Classify	What is meant . . .? What can you say about . . . ? Which is the best answer . . . ? How would you summarize . . . ? Will you state or interpret in your own words . . . ? What facts or ideas show . . . ? What is the main idea of . . . ?
3.	Application (solving problems by applying acquired knowledge, facts, techniques and rules in a different way.)	apply, build, choose, construct, develop, interview, make use of, organize, experiment with, plan, select, solve, utilize, model, and Identify	How would you solve _____ using what you have learned . . . ? How would you organize _____ to show . . . ? How would you show your understanding of . . . ? What approach would you use to . . . ? How would you apply what you learned to develop . . . ? What other way would you plan to .. . ? What would result if . . . ? Can you make use of the facts to . . . ? What elements would you choose to change . . . ? What facts would you select to show . ?

Received: 12 Feb 2020 | Revised: 15 March 2020| Accepted: 24 April 2020

			What questions would you ask in an interview with . . . ? What are the parts or features of . . . ?
4.	Analysis (examining and breaking information into parts by identifying motives or causes; making inferences and finding evidence to support generalizations.	analyze, categorize, classify, compare, contrast, discover, dissect, divide, examine, inspect, simplify, survey, take part in, test for, distinguish, list, distinction, theme, relationships, function, motive, inference, assumption, Conclusion	How is _____ related to . . . ? Why do you think . . . ? What is the theme . . . ? What motive is there . . . ? Can you list the parts . . . ? What inference can you make . . . ? What conclusions can you draw . . . ? How would you classify . . . ? How would you categorize . . . ? Can you identify the difference parts..? What evidence can you find . . . ? What is the relationship between . . . ? Can you make a distinction between.. ? What is the function of . . . ? What ideas justify . . . ?
5.	Synthesis (compiling information together in a different way by	build, choose, combine, compile, compose, construct, create, design, develop, estimate,	Can you propose an alternative . . . ? Can you invent . . . ? How would you adapt _____ to create a different . . . ? How could you change (modify) the

	combining elements in a new pattern or proposing alternative solutions.)	formulate, imagine, invent, make up, originate, plan, predict, propose, solve, solution, suppose, discuss, modify, change, original, improve, adapt, minimize, maximize, delete, theorize, elaborate, test, improve, happen, change	plot (plan) . . . ? What could be done to minimize (maximize) . . . ? What way would you design . . . ? What could be combined to improve (change) . . . ? Suppose you could _____ what would you do . . . ? How would you test . . . ? Can you formulate a theory for . . . ? Can you predict the outcome if . . . ? How would you estimate the results for . . . ? What facts can you compile . . . ? Can you construct a model that would change . . . ? Can you think of an original way for the . . . ?
6.	Evaluation (presenting and defending opinions by making	award, choose, conclude, criticize, decide, defend, determine, dispute, evaluate, judge,	Do you agree with the actions . . . ? with the outcomes . . . ? What is your opinion of . . . ? How would you prove . . . ? disprove . . . ?

judgments about information, validity of ideas or quality of work based on a set of criteria.)	justify, measure, compare, mark, rate, recommend, rule on, select, agree, interpret, explain, appraise, prioritize, opinion, ,support, importance, criteria, prove, disprove, assess, influence, perceive, value, estimate, influence, Deduct	Can you assess the value or importance of . . . ? Would it be better if . . . ? Why did they (the character) choose...? What would you recommend . . . ? How would you rate the . . . ? What would you cite to defend the actions . . . ? How would you evaluate . . . ? How could you determine . . . ? What choice would you have made . . ? What would you select . . . ? How would you prioritize . . . ? What judgment would you make about . . . ? Based on what you know, how would you explain . . . ? What information would you use to support the view . . . ? How would you justify . . . ? What data was used to make the conclusion . . . ? Why was it better that . . . ?

Note: Information included in table no 1 is adapted from the Taxonomy of educational objectives. Handbook 1: Cognitive domain (Bloom, 1956).

Bloom taxonomy shows different levels of reading skills and the ways about how to construct multiple items according to the level and needs of the students.

IV. Conclusion

It can be easily concluded that the reading comprehension test is not as easy to construct as it seems to be. The test developer has to keep many aspects into consideration while constructing test items. Different approaches in this regard can be observed as a top-down approach, bottom-up approach, and interactive treatment approach, which can help to develop a test. It is useful to include a variety of text types for reading comprehension in addition to the usual, more literary prose passages: e.g. newspaper articles, directory extracts, public notices, instructions for using appliances and machinery, timetables advertisements and maps, etc.

Researches have proven that the use of multiple-choice items as the most suitable instrument in testing reading comprehension. The text itself should always govern the types of questions that have to be constructed, particular texts may lend themselves to multiple-choice items, others to matching items, others to true-false items, others to the completion of the gaps of information in tables and many others to open-ended questions. Reading comprehension measure includes contextual factors such as type of the text, topic, theme, and purpose, and the maintaining of different subskills, recovery of explicit information, interpreting the text content, or critical evaluation.

In constructing a test for reading skills, cognitive levels of Bloom taxonomy can never be ignored, as they help the test maker to develop a test according to the level, age, and needs of the testee. Moreover, the types of tests like proficiency test, summative test, standardize test and classroom test, etc, can easily be accommodated in the light of Bloom taxonomy.

References

1. Alderson, J. C. (2000). *Assessing reading. Cambridge*: Cambridge University Press.
2. Andrich, D., Humphry, S. M., & Marais, I. (2012). Quantifying local, response dependencebetween two polytomous items using the Rasch model. *Applied Psychological Measurement*,36(4), 309–324.
3. Bloom, B. S. (1956). Taxonomy of educational objectives. *Handbook 1: Cognitive domain*. New York: McKay.
4. Campbell, J. R. (2005). A single instrument, multiple measures: Considering the use of multiple item formats to assess reading comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 347–368). Mahwah, NJ: Lawrence Erlbaum.
5. Chou, P. T.M. (2011). The effects of vocabulary knowledge and background knowledge on reading comprehension of Taiwanese EFL students. *Electronic Journal of Foreign Language Teaching,* 8(1), 108-115.Comprehension'. Language Testing 4:195-220

6. Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading* 10(3), 277–299.
7. Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement,* 27, 209–226.
8. Grabe, W. & Stoller, F.L. 2002: *Teaching and researching reading*. Harlow:Longman.
9. Grellet, F. (1981). *Developing reading skills.* Cambridge: Cambridge University Press.
10. Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New York: Routledge.
11. Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge.
12. Heaton, J. B. (1990). *Longman Handbooks for Language Teachers: Writing English Language Tests* (New ed.). London and New York: Longman Group UK Ltd.
13. Hill, C. and Parry, K. (1994) '*Assessing English language and literacy around the world*in Hill, C. and Parry, K. (eds) From Testing to Assessment. London:
14. Hill, C., & Parry, K. (1994). Models of literacy: the nature of reading tests. *From testing to assessment: English as an international language*, 7-34.
15. Hughes, A. 1989. *Testing for Language Teachers.* CUP.
16. Jafapur, A. (1987) 'The short-context technique: an alternative for testing reading comprehension'. *Language Testing* 4: 195-220
17. Jahan, K., Shah, S. K., & Ahmad, S. S. (2019). Evaluación de materialesen aulas de lenguainglesa de nivelsecundario. *DilemasContemporáneos: Educación, Política y Valores*, *6*.
18. Kintsch, W., & Yarbrough, J. C. (1982). Role of rhetorical structure in text comprehension. *Journal of Educational Psychology*, 74(6), 828–834.
19. Kobayashi, M. (2002). Method effects on reading comprehension test performance: text organization and response format. *Language Testing*, 19(2), 193–220.
20. Leslie, L., & Caldwell, J. (2009). Formal and informal measures of reading comprehension. In S. E. Israel & G. G Duffy (Eds.*), Handbook of research on reading comprehension* (pp. 403–427). New York: Routledge. Longman.
21. Nation, K., &Snowling, M. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skills. *British Journal of Educational Psychology*, 67, 359–370.
22. Pearson, P. D., & Hamm, D. N. (2005). The assessment of reading comprehension: A review of practices: Past, present, and future. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 13–70). Mahwah, NJ: Lawrence Erlbaum.
23. Pulido & Hambrick, (2008), "The virtue Circle: Modeling individual differences in L2 reading and vocabulary development" *Reading in a Foreign Language* ISSN 1539-0578Volume 20, Number 20,
24. Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language testing*, *1*(2), 147-170.

25. Spolsky, B. (1985). *The limits of authenticity in language testing. Language Testing*, 2(1), 31–40. https://doi.org/10.1177/026553228500200104
26. Van den Broek, P. (2012). Individual and developmental differences in reading comprehension: *Assessing cognitive processes and outcomes*. In J. P. Sabatini, E. R.
27. Weir, C. J. (2013). *Measured constructs: A history of Cambridge English examinations*. New York: Cambridge University Press.
28. Yamashita, J. (2003). Processes of taking a gap-filling test: Comparison of skilled and less skilled EFL readers. *Language Testing*, 20(3), 267–293.