

# ANALYZING CHARACTERISTICS OF READING TEST TASKS DESIGNED FOR UNDERGRADUATE LANGUAGE ASSESSMENT

Almas Ashraf<sup>1</sup>  
Dr. Sajida Zaki<sup>2</sup>

## ABSTRACT

*Language assessment is an integral part of any language programme. It provides the information about people's language ability that carries influential consequences for the test takers and determines their academic and professional future. Consequently, such high-stakes assessment must be effective for the test takers and the various stakeholders who use the test outcomes. Testing literature is replete with evidence that language tests are generally of poor quality and do not measure accurately what they are supposed to measure. Several research studies have proved that test tasks influence the performance of the test takers, resulting in testing experts to focus attention towards enhancing the quality of test tasks. Bachman and Palmer (1996) proposed a framework for task characteristics which includes five set of characteristics: setting, rubric, input, expected response, and relationship between input and response. Using this framework, the current study investigated the characteristics of reading test tasks designed for summative assessment at undergraduate language courses which are compulsory components of study programmes across all disciplines and with reading being a significant part of the syllabus and also of the summative assessment. Thirty exam papers within the domain of English for general purposes were collected from different public and private sector universities for the analysis of reading tasks. The findings of the study shed light on the existing weaknesses in the design of reading test tasks and their potential impact on the test takers' performance.*

**Keywords:** *Language assessment, summative assessment, reading, reading tasks, test task characteristics*

## INTRODUCTION

Language assessment plays a crucial role in any language program. The information gathered from language assessment enables us to make inference about language learners' proficiency and to inform the resulting decisions about them (Bachman, 2004). It serves as a 'common yardstick' for meaningful comparison for placement, achievement, employment as well as immigration (Hughes, 2001). If the information collected is accurate, it becomes a useful tool to evaluate the learning and teaching practices; however, if it is poorly conceived or misinterpreted, it might lead to detrimental consequences (Green, 2014). This significance is further highlighted in context like Pakistan where assessment carries high-stakes. Coombe (2009) defines high-stake assessment as one where all major decisions about learners' admission, promotion, and graduation are influenced by test scores.

The purpose of language assessment is to predict the quality of test takers' performance in real life situations through the information gathered under test conditions in order to appraise their knowledge and skills (McNamara, 2004). However, using a language is a multifaceted

---

<sup>1</sup>MS Scholar (Applied Linguistics) at Humanities Department, NED University of Engineering and Technology  
Email: almasashraf786@yahoo.com

<sup>2</sup>Professor and Chairperson, Department of Humanities, NED University of Engineering and Technology  
Email: drzaki@neduet.edu.pk

undertaking that requires multiple skills and competences (Council of Europe, 2011). Using a second language becomes a more complex phenomenon since several variables are at play in the process of second language acquisition. Learners' strive to acquire a new language with its culture and a new way of thinking (Brown, 2000). Similarly, reading in a second language also calls for multiple underlying skills and capabilities for successful comprehension. It comprises of decoding, linguistic as well as topical knowledge, and cognitive process (August, Francis, Hsu and Snow, 2006). Readers also engage in interactional processes, meta-cognitive strategies and self-monitoring (Hudson, 1996). Thus, second language learners face difficulty in comprehending the reading input (Behfrouz and Nahvi, 2013). However, the complexity of the skill cannot undermine its importance in academic and professional domains. The significance of reading further augments for students of ESL/EFL context, particularly at undergraduate level, where the medium of instruction and recommended books are in English (Shaw, 2010). To access and utilize the global reservoir of knowledge and research, Pakistani undergraduate learners need to be proficient readers. This proficiency in reading can be appraised through reading assessment as Jafarpur (2003) states a reading test determines test takers' understanding of the written texts. Therefore, to evaluate their present reading skills and to predict their future comprehension performance, reading assessment must be carried out.

Reading is a receptive skill that is manifested in overt behavior (Hughes, 2001). The direct assessment of reading is impossible since what takes place in the mind of the reader while reading cannot be observed directly. To assess reading, evidences are required to interpret what learners have comprehended. One way to collect this evidence is comprehension technique using open-ended or close-ended questions. Open-ended questions usually require production on part of the test takers in form of short responses while close-ended questions tap into the recognition ability by using MCQs, true or false, completion and reordering (Song, 2008). Fehér (2015) categorizes comprehension into hard and soft reading tasks: the former is restricted to only one correct answer or interpretation whereas the latter allows multiple interpretations on the basis of readers' personal experiences and judgment.

Reading comprehension performance is highly influenced by the text, content, and task features (Davey, 1987). Bachman (1990) called these characteristics 'test method facets' and argues that test developers have control over the design of these facets which have high impact on the performance of test takers. Additionally, the impact of these methods varies among test takers. Therefore, the test task characteristics must be considered carefully for selecting or designing a test because these tasks are the optimal source to assess test takers' performance (Behfrouz and Nahvi, 2013). The changes in the test task features might lead to change in the performance and render the assessment invalid and unreliable. In Pakistan, several researches have been carried out on assessment practices at secondary and tertiary level (Khan, 2011; Martin, 2007; Rehmani, 2007; Qureshi, Shirazi & Wasim, 2007; Raza, 2009). However, most of them discuss the existing gaps and weaknesses in the examination system on the whole. There is a need for in-depth research targeted at assessment of specific skills and the methods through which they are assessed in order to have stronger belief and confidence in the decisions made on the basis of these assessment practices.

Therefore, the aim of the present study is to analyze the characteristics of the test tasks designed for reading assessment for undergraduate English for general purposes course in order to examine its alignment with the framework of Bachman and Palmer (1996). The scope of the present study is restricted to reading test tasks and does not include test tasks

for any other language skills. The courses selected for the analysis of reading test tasks fall into the category of English for general purposes taught in undergraduate programs. The study has also its limitation. As the current research collected examination papers only, it does not include the analysis of setting. The characteristics of the setting in the framework comprise of three components: physical setting in which the authors discussed location, noise level, temperature, humidity, seating conditions, lighting and familiarity of material and equipment used for test; participants which involves test takers and administrators; and time of task which addresses whether the test takers are fresh or fatigued at the time of the test.

## LITERATURE REVIEW

Reading has been defined in terms of process and product. The process refers to the interaction of text and a reader and the resulting progression of thinking and meaning-construction. These internal and silent processes are dynamic in nature and may vary for the readers depending on the text, time and the purpose of reading. Researches on reading process are focused on readers' eye movement, reading aloud, miscue analysis, think aloud protocol and verbal retrospection. Alternatively, product of reading refers to the understanding that the readers reach as a result of these processes (Alderson, 2000). The understanding of a reading text involves identifying meaning of and relationships among words using the prior knowledge of grammar as well as a more active assimilation of the text information and previous background knowledge (Montgomery, Durant, Fabb, Furniss & Mills, 2007). Multiple perspectives have been presented to understand the construct of reading competence. Cognitive perspective highlights reading competence as decoding, meaning construction and synthesis of new and previous knowledge whereas development perspective emphasizes the sequential yet interdependent processes of decoding and comprehension. On the other hand, the reading gear theory incorporates the purpose of reading along with these two components in the definition of reading competence (Koda, 2004).

There are various reasons that lead to poor performance on reading test; they are inaccurate word reading, lack of reading fluency, lack of interest in reading, weak vocabulary or limited background knowledge, and varying socio-economical background of test takers. Similarly, test tasks affect the reading performance of the test takers. Carroll (1993) defines tasks as an activity which allows a person to engage in an appropriate setting to achieve specific objectives. Salmani-Nodoushan (2003) explored the effects of text familiarity, task types and language proficiency on the reading comprehension and concluded that all three variables are influential factors that affect the reading performance of students. Aghajani, Motahari and Qahraman, (2013) conducted a similar study to explore the relationship between text familiarity and task type with the reading performance and found that both the variables significantly affect the reading comprehension. Students who were familiar with the content of the test performed better than their counterparts. Similarly, reading comprehension and performance differed on the basis of different task types. Alderson, Clapham and Wall (2010) have drawn attention to the 'method effect' explaining that the test technique will strongly affect students' performance on the test; therefore, test developers should consider what will be the effect of test task. In this respect, Education policy of Pakistan also emphasizes that test takers' skills should be assessed through multiple techniques considering the purpose and objectives of the course (Ministry of Education, 2009)

Valid and reliable test construction for reading ability was once considered an easier task compared to the construction of writing and speaking (Hughes, 2001). Grabe and Jiang (2014) summarize the 'intriguing history' of reading comprehension assessment. Prior to 20th century, assessment of reading was focused on literary and cultural interpretation leading to subjective measurement. However, in the decades of 1960's and 1970's, objective testing was encouraged and tests like TOEFL and IELTS emerged. In the succeeding years, the limited scope of objective testing to infer the reading skills of individuals was challenged and the emphasis shifted to communicative and integrative assessment of reading. The end of the century witnessed cognitive research and the resulting characterization of reading sub-skills. From 20th century onwards, there is a growing need of reading and comprehending large amount of information and its use for academic and professional purposes.

It has been a common understanding that language assessment is carried out to make inference about test takers ability to perform in real life situations through their performance in specific test conditions. However, there is a need of a model or a framework to align the task characteristics of real life usage with the test settings (Bachman and Palmer, 1996). In this regard, language assessment models accomplish dual purposes for the validation of any test: it serves as a framework for blueprint or test specifications (Alderson, Clapham and Wall, 1995) as well as a mechanism to ensure the alignment of the test construct and the inference based on that test (Messick, 1989).

Various models and frameworks have been presented in the history of language testing according to the theories of language acquisition and language use. Lado (1961) proposed his model in which he divided language into skills and components. His model followed discrete-point testing approach, although he acknowledged that these skill and especially elements are not used in isolation. Lado's model was a product behaviorist theory of language acquisition where language is acquired through habit formation and drills. Oller in 1979 challenged this approach and prompted integrative and pragmatic approach to language testing. He saw cloze technique as an embodiment of integrative testing. He also proposed 'Unitary Competence Hypothesis' which argued that all language tests measure a single underlying construct, i.e. language ability, however, it soon fell out of favour (Green, 2014). Later on, building on the previous communicative models of Canale and Swain (1980) and others, Bachman (1990; Bachman and Palmer 1996; 2010) proposed a language model that treats language knowledge as discrete yet interdependent competences. Language of knowledge, according to them, comprises of organizational competence (grammatical and textual competence) and pragmatic competence (functional and sociolinguistic competence). Although, this model has also been criticized for lack of explanation of its contribution and dynamics in communication, it has been agreed upon that language ability is made up of several components and their assessment should be conceptualized in terms of its purposes. Considering this impact of test tasks on performance, Bachman and Palmer (1996) proposed a framework, based on Bachman (1990), for test task characteristics. The framework consists of five aspects of a task and its set of features, i. e. setting, rubrics, input, expected response, and relationship between input and expected response. They define that the purpose of the framework is to serve as a foundation for development and use of language test. By development and use they mean: description of target language use (TLU) tasks to design language test tasks, description of various test tasks to ensure comparison and reliability, and comparison of TLU and test tasks to judge the authenticity of the test.

They discussed this framework for the design and construction of language tests yet the flexible and adaptive nature of the framework can be helpful in empirical investigation and other related researches on the already existing tests (Behfroz and Nahvi, 2013). These task characteristics refer to both the test as well as TLU setting. The first set of characteristics, setting, involves the physical circumstance in which language test or language use takes place. These characteristics include physical characteristics, participants and time of task. Physical characteristics comprise of the location, noise level, temperature, humidity, seating condition, and familiarity with the equipment and material. These characteristics include all these features which are part of the physical circumstances of the situation including weather and lighting. By participants, they mean all the concerned people who are involved in language test or use task. For language use, all the people engaged in the communication process with their different roles form the participants whereas for language test, test takers and all the concerned people in test administration will be considered participants. Their mutual relationship and familiarity will also be considered. Time of the task simply refers to the time frame in which the test or use takes place; time is an influential factor for language performance. Rubrics of the test task characteristics include the structure of the task and instructions on how to accomplish a task. This set of characteristics is highly significant for the language test setting and, therefore, must be made explicit and clear. Along with structure and instructions, this set of features also contains time allotment and scoring method. Instructions involve language, and channel of presentation, and procedures to be followed, whereas structure of the task contains information about the number, salience, sequence and relative importance of the tasks. Time allotment is the duration specified for individual tasks as well as the entire test; they discuss speeded and power test based on the time allotted to the test takers to complete the tasks. The last characteristic of rubric, scoring method, refers to the method of evaluation of the responses that includes criteria for correctness, procedures for scoring response, and explicitness of criteria and procedure.

Input is anything that is provided to the test takers or language users as a prompt or stimulus to perform certain tasks. Input is discussed in terms of format and language. The format includes channel of presentation, form, language, length, type of input, degree of speededness and vehicle. On the other hand, the language refers to language characteristics - organizational and pragmatic - and topical characteristics. Organizational and pragmatic characteristics are further classified into grammatical and textual characteristics and functional and sociolinguistic characteristics respectively.

Expected response is differentiated from the actual response that is presented by test takers as test takers or language users are people and they might not understand or be reluctant to respond in a particular way. Therefore, the actual responses may or may not be consistent with the expected responses. Expected response deals with format, types of response (selected, short or extended), degree of speededness and language. These characteristics are similar to the characteristics of input. The last set of characteristics deals with the relationship between input and response. It discusses reactivity, scope of relationship, and the directness of this relationship. While reactivity involves the degree of interaction between input and response, the scope refers to the amount of processing input to produce response and directness of relationship is the extent to which the expected response relies on the input.

## METHODOLOGY

The current study employs qualitative paradigm of research to analyze the test tasks characteristics of reading skills for undergraduate courses of English for general purposes. Qualitative research collects, analyzes and interprets non-numerical data in order to obtain insight that evolves with the understanding of the context (Gay, Mills & Airasian, 2011). Through this approach, it aims to answer the following research question:

“ What are the characteristics of the reading test tasks designed for undergraduate English courses?

To answer this question, 30 examination papers were collected from several public and private universities which were used for the summative assessment in these universities. It was ensured that only papers for English for general purpose courses that assessed reading skill were collected to maintain reliability of the findings and the courses specifically designed for academic writing or speaking were excluded from the sample (see Table 1 for key features of the paper). While collecting these summative assessment papers, the researchers directly approached the teachers at these universities who had developed and administered these papers. They were requested to share the papers which were already open and public documents as the institute's policy allowed students to carry the papers with them after the exams, and the same were available at the libraries or book shops or photocopiers. The teachers' informed consent, assurance of complete confidentiality and anonymity of the teachers as well as their universities, withdrawal from the study and right to know the findings were the key ethical standards focused by the investigators. Moreover, the teachers were asked to remove the institutions identity that appeared on the question paper to further ensure that the institution and the teacher's details do not become public while the study was being conducted. However, the coding of these papers enabled in classifying these papers in terms of general features that appear in Table 1.

---

n	Number of Question Papers	30
n	Number of Reading Questions	49
n	Weightage of Reading Questions(Range)	5-20 marks
n	Distribution of Question Papers with respect to:	9 Professional, 7 General
o	University Type (General / Professional)	
o	University Type (Public / Private)	5 Public,11 General
n	Undergraduate Study Programmes	Bachelor's in Engineering (several disciplines within engineering), Bachelor's in Architecture, Bachelors of Science (several disciplines within basic science, humanities, and social science)

---

## DATA ANALYSIS AND FINDINGS

The collected papers were then analyzed using content analysis. Neuman (2015) defines content analysis as an analysis of the content of the text through a coding system. He also states the unit of analysis may vary from a word to a character. For the present study, reading

test task is selected as a unit of analysis. The categories for coding were predetermined from the framework which consists of five set of characteristics (see appendix). However, as already indicated in the study limitations, the study employed four criteria only excluding the characteristics of the setting. The reading test tasks in the collected samples were analyzed on the basis of rubrics, input, expected response and relationship between input and expected response.

The content analysis of the papers shows similarity among the reading assessment tasks across different universities with little variations in some of the aspects. The teachers set the reading tasks in similar fashion using the same techniques to elicit understanding of the text with few exceptions. It was surprising to note that four out of thirty papers which targeted English for general purposes did not assess reading at all. No single question was found in these four papers which tapped into any of the major or minor sub-skill of reading. The assessment of reading was carried out mostly by setting tasks for identification of topic sentences or thesis statement, understanding explicit information given in the text, inference, summary of extended discourse, and guessing words meaning in context. On the other hand, there were few instances where the teachers targeted, along with these macro- and micro-skills, identification of pronoun referents, understanding author's writing style, evaluation of the main idea with reference to their personal experience, and error correction. Among all the papers, only single cases were found for recognition of genre, identification of spoken features and scanning.

The findings of the study are discussed in terms of four major set of characteristics proposed by Bachman and Palmer in their framework (see Table 3 for consolidated findings).

#### **Findings: Characteristics of Rubrics**

This set of characteristics includes the instructions, structure, time allotment and scoring method of the reading test tasks. The instructions were given in the target language in written form. Oral instructions might have been provided to the test takers depending on the invigilators and the need of the students, however, it is out of the scope of the current study. Procedures or specifications of instructions were found to be uniform across all the tasks: there were brief instructions to 'attempt all questions' or 'read the given passage carefully and answer the following questions'; no instruction followed any example of expected responses; and general instructions, or rather directives, were given on individual tasks.

Regarding the structure of the tasks, it was found that reading tasks varied in number in the papers: nine papers had only one reading task in the entire exam paper, the other nine had two reading tasks, four papers had three reading tasks, three papers had four and one paper had five reading questions. These tasks were not distinguished from other tasks in any way apart from the fact that they were focused on assessment of reading while the rest of the tasks targeted other skills. As far as relative importance of these reading tasks is concerned, it was found that only three reading tasks (P2, P3, P4) were given as 'compulsory questions' emphasizing the relative importance of reading whereas in other cases they were presented with no relative emphasis since the test takers were expected to attempt all the given tasks. It was also found that no restrictions were imposed on the order of the tasks except for one paper (P25) in which it was clearly mentioned that test takers should 'solve all questions in

sequential order given in the paper'. These reading tasks were further divided into sub-parts, usually called test items, to elicit information on the different parts of the given input. The number of these sub-parts ranged from 4 to 10 items where nine tasks were followed by 5 test items, six tasks by 4 test items, six tasks by 8 test items, and five tasks by 10 test items. With regards to time allotment to individual task, no time distribution was specified, although the duration of the complete test was clearly mentioned on the paper which ranged from 2 to 3 hours depending on the allocation of the marks for the final exams. Similarly, scoring method was not mentioned in any of the test tasks explicitly except in two tasks (P29 and P30) where it was given vaguely: 'spelling, grammar and punctuation mistakes may be penalized'. Apart from this, no information on the evaluation of the responses were mentioned in any of the tasks, although the marks distribution was clearly spelled out for each task and its sub-parts. With regards to objectivity and subjectivity of scoring, the type of the test determines the approach. If the task is selected or limited (having only one correct answer) the examiner may use an objective scoring key but the extended response requires subjective scoring. As the present study analyzed the exam papers only and there was no statement on the use of scoring key or rating scale, it is limited in the analysis of scoring procedures.

### **Findings: Characteristics of Input**

As reading involves the decoding of written and visual text, all the reading test tasks included written input in the target language, i. e. English. Additionally, the reading input for comprehension as well as summary fell into the category of extended discourse. Five passages among all were relatively lengthier than other whereas sixteen passages were of medium difficulty and six were short passages. These extended stretches of discourse were followed by open and close-ended items for comprehension tasks. For summary tasks, only three tasks were found which provided guided input in the form of completion tasks; the test takers were required to complete a summary with blanks, a table, and a flow chart (P23 and P24). Moreover, it was found that the same passage was used for comprehension as well as summary in all but four test tasks (P8, P16, P22, P23 and P24).

Degree of speededness, or the rate of processing the input, was left unspecified and no such input was provided as the papers are distributed on the specific point in time and collected after the specified duration of the test. The time spent on individual tasks is not recorded or restricted in usual practices. However, there was one exception (P14) which instructed the test takers to complete the task in the quickest possible time in the following words: 'scan the travel brochure and find the answers to the following questions as quickly as you can'. How this quickness was recorded and ensured is left unanswered.

The analysis of language of the input revealed that mostly the selected reading input was adopted from other sources; one of the teachers mentioned the source of the text in the paper (P3). These reproduced pieces of text followed all the conventions of grammatical and textual features including vocabulary syntax, rhetorical organization and cohesion. However, the input for two reading tasks (P15, P22) that targeted learners' judgment of errors was found to have errors of clutter, punctuation, cohesion, fragmentation, run-on sentences, comma splices, and faulty parallel structure. Similarly, pragmatic characteristics of the input that include functional and sociolinguistic features were also analyzed. It was found that the selected input was mostly ideational in its functional approach. Nevertheless, the style of writing varied from academic register to natural style using the standard variety of English.



Analyzing the topical characteristics, it was revealed that the type of information provided in the test input ranged from personal to technical topics (see Table 2). However, it was observed that cultural topic outweighed all other domains of topical knowledge as majority of the topic is cultural and includes topical information not only about the culture of the context of the current study but also of different other parts of the world.

**Table 2:**  
**Topics selected for reading input**

Type of Topic Information	Topics
Personal (12)	Chocolate, Money, Family, Travel to Europe, Summer vacations, Envy, Grandparents, Marriage, Parents, Jobs, Driving and age, Man who survived poverty.
Cultural (23)	Making Pizza, Human rights, Tropical lake, Euthanasia, Marrying an expat, Bollywood, Cultural differences, Weather, Parks, Farming, Population Growth, Tornadoes in Chicago, Crimes at college campuses, Helping each other, Abraham Lincoln, Dogs in America, Use of garlic, Boys and girls, Horace Mann, Young rebel, Persian household, Jammu Kashmir, Volcano eruption.
Academic (6)	Plagiarism and cheating, history of television, note-taking, Human cell, Pollution, Hedgehogs.
Technical (7)	American aviation service, HRM policies, Professions, Travel agency brochure, Galaxies, Law of dynamics, advertisements.

### Findings: Characteristics of Expected Response

Bachman and Palmer (1996) outlined format, type of response and language of expected response but, surprisingly, no explicit instructions were given with reference to these features for the selected reading test tasks. Nevertheless, the test takers could infer from the test design what kind of responses were required that might lead to variation in interpretations and adverse effects on performance. As far as format is concerned that includes channel, form, language, length, and degree of speededness, the test takers are aware that they are expected to write answers in the target language. However, for the process of input and planning of response, no time allocation or degree of speed was specified for individual tasks and test takers were allowed to invest as much time as they required for producing responses for different tasks provided they finish the paper by the specified time. As the rate of processing information and other related cognitive process vary in individuals, this freedom of choice may be interpreted as a threat to the validity of the inferences made by such assessment.

The length of responses was determined by the type of response expected from the test takers. The response length and response type in the selected papers show that most of the tasks elicited only one type of response. In eighteen reading tasks only extended responses were expected, mostly short answers and summary. Eleven tasks employed only selected response to assess reading ability where test takers were required to select a response from the given choice. Such practices limit learners' opportunities to demonstrate their comprehension skills to the fullest. Nine tasks were designed to elicit only short responses which allowed test takers to produce a response but it ranged from a word to a phrase only. Ten out of thirty

papers showed a combination of these types of responses where test takers were not only required to select the correct response from the given choices but also to produce a limited or an extended response; thus, giving more opportunities to them to display their reading skill comprehensively.

For selected and short responses, the test takers were not expected to produce any text beyond the given input. Resultantly, no information was given about the language of the expected response. But no statement was made on the language of response even for the tasks where test takers were required to write a short answer or a summary. They were not provided with any specification of the grammar, vocabulary, syntax, cohesion or rhetoric. Additionally, no instructions were found during the papers analysis that spelled out the functional or topical features of the expected response. Since expected responses were directly related to the input, the test takers were restricted to the topic of the given input. The only exception to this common trend was the tasks that were set for error correction (P15, P22); these tasks instructed the test takers to rewrite the given text after removing various grammatical and textual errors. But it was only limited to language features, pragmatic and topical characteristics still did not find any place in the specification of the expected response. Similarly, only two tasks (P6 and P1) were found where responses were based on their personal experience.

#### **Findings: Relationship between Input and Expected Response**

The relationship between input and the expected response is discussed in terms of reactivity, scope and directness of relationship. Unlike speaking, reading is tested without any feedback from the testers or administrators, therefore, the reactivity of the all the test tasks was non-reciprocal. The tasks were designed to elicit learners' performance on the task once with no provision of feedback or improvement. In addition to this, majority of the reading tasks, nineteen, utilized both broad as well as narrow scope to assess reading comprehension. On the other hand, five tasks (P11, P12, P15, P16 and P28) targeted only broad scope understanding and two tasks (P25, P26) aimed for narrow scope only. Broad scope was used for comprehension of extended discourse, writing summary, and reordering jumbled sentences to make a coherent paragraph whereas narrow scope included inference of selected words meaning, identification of pronoun referents, and answering explicit questions.

The common trend for reading assessment is to select a passage and design several open and close-ended questions that require reproduction or rephrasing of the information given in the input. Consequently, the relationship of input and expected response was direct in thirty nine tasks where the expected response was based on the given input. Six tasks were found to have both direct and indirect relationship resulting in better and more valid inference of test takers' reading ability. However, there were only two tasks (P12 and P14) with indirect relationship of input and expected response; the test takers were required to produce moral of the given narratives which was not explicitly given in the input.

Table 3: Consolidated findings of analysis of reading test tasks

<b>Consolidated Findings</b>	
Characteristics of rubrics	<ul style="list-style-type: none"> <li>n Brief and general instructions in the target language without any example of expected response</li> <li>n Number of reading tasks in a paper ranged from 1 to 4 with 4 to 10 items/questions where most of the papers assessed reading skill through a single task only.</li> <li>n Reading tasks were compulsory to attempt even when there was a choice of questions in the papers; there was no restriction on the order of the tasks except in one paper.</li> <li>n No time distribution for individual tasks was specified.</li> <li>n No information on scoring criteria, procedure or rating scales was provided.</li> </ul>
Characteristics of input	<ul style="list-style-type: none"> <li>n Written extended discourse in target language (English);</li> <li>n Few tasks provided guided input for task completion. Mostly same passage was used to assess comprehension and summarizing skills.</li> <li>n Adapted text which followed all the conventions of grammar and rhetoric except the passages in which the test takers were required to identify and correct errors.</li> <li>n Mostly ideational in its function; writing style varied from academic register to natural style.</li> <li>n A range of topics used but cultural and personal topics outweighed academic and technical ones.</li> <li>n No specification was given about the planning or processing time except one scanning task.</li> </ul>
Characteristics of expected response	<ul style="list-style-type: none"> <li>n Written responses in target language (English).</li> <li>n No explicit statement regarding length or processing input was made; test takers were required to infer the information through the type of the tasks.</li> <li>n Mostly one type of response (either short or limited or extended) was elicited thus restricting test takers' to demonstrate their reading skills fully. Few papers aimed at combination of these types.</li> <li>n Linguistic or pragmatic features of the expected responses were not spelled out in any of the tasks except the tasks for error correction.</li> <li>n The topics of the responses were directly related to the topics of the input. Instructions on functional features were not given, however, linguistics features were only mentioned in error correction tasks.</li> </ul>
Characteristics of relationship between input and expected response	<ul style="list-style-type: none"> <li>n Non-reciprocal relation between input and output with no feedback.</li> <li>n A combination of broad and narrow scope was employed to assess test takers' understanding at local as well as global level.</li> <li>n Direct relationship between input and response in most cases with few exceptions of indirect relationship or the combination of the two.</li> </ul>

## DISCUSSION

The findings of the present study show that assessment of reading is very limited in its nature, range and scope. The overall examination of the papers revealed that writing tasks dominate the exams and reading tasks have little importance relatively. In most of the cases, there was only one task targeted at reading which does not reflect the test takers' reading ability comprehensively and completely. Moreover, the test tasks in the present analysis were found to be similar in their approach and design regardless of the differences in the courses which suggest that perhaps the test tasks are easily predictable and are, therefore, no more valid. These findings ignore the basic principles of assessment as well as the recommendations of testing experts who suggested to consider the effect of test tasks (Alderson et al., 2010) as well as the instruction towards incorporating variety of techniques for assessing test takers (Ministry of Education, 2009) .

Although these tasks employed all three types of responses - selected, limited and extended - but there is a vast choice of test tasks among these three major types. The analysis revealed that mostly 'MCQs' and 'True or False' were used as selected responses, identification of topic sentences, main idea and errors as limited responses and summary and short answers for extended responses. Alderson (2000) reached similar conclusions that MCQs are the most commonly technique used for the assessment of reading although it has been challenged as an indirect assessment of reading ability. The purpose of reading assessment is to judge the test takers' present competence as well as to predict their future endeavors in reading, however, these limited choices of test types might not reflect the test takers' reading ability truly. Limited tests means less evidence is collected about their reading abilities (Fulcher and Davidson, 2007). Similarly, responses to multiple tasks should not be mutually dependent so that each response elicits unique information about the test takers. Nevertheless, such interdependence is irrelevant here as reading was assessed through single task in most cases. With regards to the reading sub-skills, comprehension of the explicit meaning of the text outweighed all other types of reading sub-skills in the papers regardless of the fact that students at undergraduate are not only require to read the lines but also to read between and beyond the lines. Sherman (1997) reinforces the same constraints and argues that comprehension questions do not represent TLU domain. This, again, indicates toward the limited and partial assessment of reading.

In addition to this, the current reading test tasks do not reflect the modern or alternative techniques to test reading; reading is assessed through the old traditional methods. This traditional approach is not able to capture the interactive and complex process taking place between the text and the test takers (Heinz, 2004). Khan (2011) discusses the effects of using tradition approach with reference to the context of the present study that it has adverse effects on the quality of education as the assessment carries high-stakes consequences for the test takers. Educational systems where assessment holds and influential status tend to have its backwash effect on learning and teaching itself. The classroom practices are determined by the test contents and past papers becomes the curriculum. Similar impact of examinations has been observed in Pakistani classrooms (Rehmani, 2007). Consequently, these traditional practices of reading assessment would drive the learning and teaching practices in the classroom resulting in limited learning and reading development.

The significance of evaluation criteria for reading increases in extended responses as writing is involved to demonstrate comprehension and inference. Test takers need to be aware of the criterion of scoring and relative importance of correctness in reading and writing. As Hughes (2001) points out that test techniques should not interfere with the reading process because some test takers might comprehend perfectly but might face difficulties in the written demonstration of that understanding. He suggests that spelling, grammar and punctuation mistakes should be overlooked and not penalized while scoring a reading test if a test taker completes the targeted task successfully. His recommendations to use close-ended questions like MCQs and true or false might not be applicable completely as it has already been discussed that reliance on only selected and limited responses to elicit reading performance is not enough. Therefore, with the combination of selected, limited and expected response, evaluation criteria for the reading tasks must be explicitly stated.

Absence of explicit and focused instructions might lead to variation in the interpretation of the given tasks resulting in varied performance of the test takers. This variation of interpretation may affect the validity of the test results. On the other hand, it was also observed that the topics used for reading input mostly involved Western culture or culture from other parts of the world. The test takers of Pakistan might not be aware of certain cultural and geographical aspects of the text, e. g. the volcano eruption or tornadoes in Chicago. The selection of such texts endorses Fulcher and Davidson's (2007) stance that test setters tend to select input that is readily available or feel relevant intuitively.

## CONCLUSION

The present study aimed at analyzing the test task characteristics designed to assess reading skills of undergraduate students of Karachi. The results show that the current reading test tasks are not comprehensive enough to reflect test takers reading skills and abilities. They are limited in its number, type, specifications and scope. This implies that the use of new and multiple techniques might result in better and stronger inferences about test takers' competence. The study is limited to the assessment practices of Karachi so the results can only be generalized to the test setters of the city, particularly, at undergraduate level and only for the courses of English for general purposes. However, the findings of the study carry implications for various stakeholders involved in the process of education and assessment since it highlights the existing weakness in the design of reading test tasks. Reading is one of the two skills assessed at the end of the term to test students' achievement in language and to decide about their promotion or graduation; therefore, it needs to be as valid and reliable as possible. Further studies can be carried out in the field to study test administration including the characteristics of setting and evaluation procedures in order to obtain a fuller and more comprehensive picture of the investigated phenomenon. Additionally, the framework can be used to design new reading tests as well as improve already existing ones. The scope of the current study can also be expanded by incorporating all four skills in the analysis of test tasks.

## REFERENCES

- Aghajani, M., Motahari, M. & Qahraman, V. (2013). The effect of text familiarity and reading tasks on ESP test performance. *Journal of Emerging Trends in Computing and Information Sciences*. 4(3), 296-302.
- Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Alderson, J. C., Clapham, C. & Wall, D. (2010). *Language test construction and evaluation*. Cambridge University Press, South Asian Edition, New Delhi: Swan Press.
- August, D., Francis, D. J., Hsu, H. A. & Snow, C. E. (2006). Assessing reading comprehension in bilinguals. *The Elementary School Journal*, 107(2), 221-238.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

- Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.
- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice: designing and developing useful language tests*. Oxford: Oxford University Press.
- Behfrouz, B. & Nahvi, E. (2013). The effects of task characteristics on IELTS reading performance. *Open Journal of Modern Linguistics*, 3(1), 30-39.
- Brown, H. D. (2000). *Principles of language learning and teaching*. NY: Longman.
- Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analysis studies*. Cambridge: Cambridge University Press.
- Coombe, C. (2009). Washback and the impact of high-stakes tests on teaching and learning. In Mansoor, S., Sikandar, A., Hussain, N. & Ahsan, N. M. (Eds.). *Emerging issues in TEFL: Challenges for Asia* (298-307). Oxford: Oxford University Press
- Council of Europe (2011). *Manual for language test development and examining*. Strasbourg: Council of Europe.
- Fehér, J. (2015). From everyday activities to creative tasks. In Maley, A. & Peachey, N. (Eds.). *Creativity in the English language classroom* (64-72). London: British Council.
- Fulcher, G. & Davidson, F. (2007). *Language testing and assessment: An advanced resourcebook*. NY: Routledge
- Gay, L. R., Mills, G. E., & Airasian, P. W. (2011). *Education Research: Competencies for Analysis and Application*. South Asia: Pearson Education Ltd.
- Grabe, W. & Jiang, X. (2014). Assessing reading. In Kunnan, A. J. (Ed.). *The companion to language assessment*. DOI: 10.1002/9781118411360.wbcla060
- Green, A. (2014). *Exploring language assessment and testing: Language in action*. London: Routledge.
- Heinz, P. (2004). Towards enhanced second language reading comprehension assessment: Computerized versus manual scoring of written recall protocols. *Reading in a Foreign Language* 16(2). Retrieved from <http://nflrc.hawaii.edu/rfl/October2004/heinz/heinz.html>
- Hudson, T. (1996). *Assessing second language reading from a communicative competence perspective: Relevance for TOEFL 2000*. New Jersey: Educational Testing Service

- Hughes, A. (2001). *Testing for language teachers*. New York: Cambridge University Press.
- Jafarpur, A. (2003). Is the test constructor a facet?. *Language Testing*, 20(1), 57-87.
- Khan, I. (2011). Reading assessment techniques among selected secondary school teachers in Pakistan: Current trends and practices. *International Journal of New Trends in Education and their Implications*, 2(4), 58-75.
- Koda, K. (2004). *Insights into second language reading: A cross-linguistics approach*. NY: Cambridge University Press.
- Lado, R. (1961). *Language testing*. London: Longman.
- Martin, T. (2007). Factors that inhibit middle and secondary school teachers in Pakistan from adopting a student-centered approach towards the teaching and learning process. In Ali, S. & Rizvi, M. (Eds.). *Quality in education: Teaching and leadership in challenging times (677-698)*. Proceedings of the International Conference. Karachi: Aga Khan University Institute for Educational Development.
- McNamara, T. (2004). Language testing. In Davies, A. & Elder, C. (Eds.). *The handbook of applied linguistics (763-783)*. Oxford: Blackwell.
- Messick, S. (1989). Validity. In Linn, R. L. (Ed). *Educational Measurement (13-103)*. New York: Macmillan.
- Ministry of Education. (2009). *National education policy*. Government of Pakistan. Islamabad.
- Montgomery, M., Durant, A., Fabb, N., Furniss, T. & Mills, S. (2007). *Ways of reading: Advanced reading skills for students*. NY: Routledge.
- Neuman, W. L. (2015). *Social research methods qualitative and quantitative approaches*. New Delhi: Dorling Kindersley India Pvt. Ltd.
- Qureshi, M. A., Shirazi, R. A. & Wasim, M. P. (2007) Perspectives and prospects of commencing new education policy (NEP) of Pakistan: A review of conference. *Indus Journal of Management & Social Sciences*, 1(2), 176-176.
- Raza, W. (2009). English language testing in higher education of Pakistan. *Market Forces*. 180-103.
- Rehmani, A. (2007). Teacher education in Pakistan with particular reference to teachers' conceptions of teaching. In Ali, S. & Rizvi, M. (Eds.). *Quality in education: Teaching and leadership in challenging times (495-524)*. Proceedings of the International Conference. Karachi: Aga Khan University Institute for Educational Development.
- Salmani-Nodoushan, M. A. (2003). Text familiarity, reading tasks, and ESP test performance: A study on Iranian LEP and non-LEP university students. *The Reading Matrix*, 3(1),

1-14.

Shaw, P. (2010). Research reports in academic and industrial research. In Ruiz-Garrido, M. F., Palmer-Silveira, J. C. & Fortanet-Gomez, I. (Eds.). English for professional and academic purposes. NY: Rodopi.

Sherman, J. (1997). The effect of question preview in listening comprehension tests. *Language Testing*, 14, 185-213.

Song, M. (2008). Do divisible subskills exist in second language (L2) comprehension? A structural equation modeling approach. *Language Testing*, 25(4), 435-464.